

59847

STIC-Biotech/Ch mLib

From: Chan, Christina
Sent: Thursday, February 07, 2002 11:19 AM
T : Chen, Shin-Lin
Cc: STIC-Biotech/ChemLib
Subject: RE: Rush sequence search: 09/391,606

CRFE

Please rush. Thanks Chris

-----Original Message-----

Fr m: Chen, Shin-Lin
Sent: Wednesday, February 06, 2002 6:03 PM
To: Chan, Christina
Cc: STIC-Biotech/ChemLib
Subject: Rush sequence search: 09/391,606

09/391,606 is an amended case due next bi-week and require sequence search. Could you approve the following Rush Sequence Search for 09/391,606? Thanks!

SEQ ID Nos. 1-4, 7-9 and 12-16.

Shin-Lin Chen

AU 1632
CM1 12A15
Mail Box # 12E12
(703)305-1678

POINT OF CONTACT:
BARB O'BRYEN
TECH. INFORMATION SPECIALIST
STIC CM1 ~~12014~~ 308-4291

6A05

TYPE OF SEARCH:		VENDOR/COST(where applic.)
Searcher: <u>Barb</u>	NA Sequences: _____	STN: _____
Phone: _____	AA Sequences: _____	DIALOG: _____
Location: _____	Structures: _____	Questel/Orbit: _____
Date Picked Up: _____	Bibliographic: _____	DRLink: _____
Date Completed: <u>2-8-02</u>	Litigation: _____	Lexis/Nexis: _____
Searcher Prep/Review: _____	Full text: _____	Sequence Sys.: _____
Clerical: _____	Patent Family: _____	WWW/Internet: _____
Online time: _____	Other: _____	Other (specify): _____

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:34:38 ; Search time 137.02 Seconds
(without alignments)
212.997 Million cell updates/sec

Title: US-09-391-606-15
Perfect score: 2103
Sequence: 1 MLPVGNPSPDLSLLIDTWE.....QKLISEEDLNSAVDHHHHH 394

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID22/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID22/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID22/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID22/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID22/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID22/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID22/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID22/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID22/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID22/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID22/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID22/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID22/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID22/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID22/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID22/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID22/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID22/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID22/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID22/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID22/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID22/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1925	91.5	391	20 AAY53319	Chlamydia pneumoni
2	1647	78.3	343	20 AAY56771	C. trachomatis ser
3	1534.5	73.0	389	20 AAW98188	Chlamydia psittaci
4	1408	67.0	402	20 AAW98189	Chlamydia psittaci
5	1408	67.0	525	21 AAB13645	C. pneumoniae sero
6	1408	67.0	525	22 AAG83213	protein encoded by
7	1407	66.9	402	20 AAW98187	Chlamydia psittaci
8	1356.5	64.5	387	20 AAY56767	C. trachomatis ser
9	1349	64.1	404	20 AAY56768	C. trachomatis ser
10	1345.5	64.0	376	19 AAW76366	C. trachomatis JMI
11	1338.5	63.6	397	20 AAY56766	C. trachomatis ser

12	1336.5	63.6	397	20 AAY56763	C. trachomatis ser
13	1326	63.1	396	20 AAY56764	C. trachomatis ser
14	1321.5	62.8	397	20 AAY56765	C. trachomatis ser
15	1320.5	62.8	376	19 AAW76363	C. trachomatis JMI
16	1313.5	62.5	393	20 AAY56757	C. trachomatis ser
17	1312.5	62.4	393	20 AAY56759	C. trachomatis ser
18	1308.5	62.2	372	19 AAW76365	C. trachomatis JMI
19	1306.5	62.2	393	22 AAE06646	Chlamydia trachoma
20	1306.5	62.1	374	19 AAW76364	C. trachomatis JMI
21	1305.5	62.1	395	20 AAY56762	C. trachomatis ser
22	1300	61.8	373	19 AAW76362	C. trachomatis JMI
23	1295	61.6	394	18 AAW15149	Chlamydia trachoma
24	1295	61.6	394	19 AAW73141	Chlamydia trachoma
25	1295	61.6	394	21 AAY81268	Chlamydia trachoma
26	1294	61.5	394	19 AAW57775	Chlamydia trachoma
27	1293	61.5	394	20 AAY56761	C. trachomatis ser
28	1289	61.3	392	20 AAY56760	C. trachomatis ser
29	1287	61.2	394	20 AAY56758	C. trachomatis ser
30	1284	61.1	394	20 AAY37494	Chlamydia trachoma
31	1280	60.9	394	7 AAP60004	Sequence of a majo
32	1264.5	60.1	343	20 AAY56769	C. trachomatis ser
33	1248	59.3	457	21 AAY82388	C. trachomatis MOM
34	1214.5	57.8	484	21 AAY82389	C. trachomatis MOM
35	1182	56.2	356	20 AAY56770	C. trachomatis ser
36	1134.5	53.9	514	21 AAY82391	C. trachomatis MOM
37	1085.5	51.6	454	21 AAY82390	C. trachomatis MOM
38	1053	50.1	349	21 AAY82392	C. trachomatis MOM
39	913	43.4	228	20 AAW98184	Chlamydia psittaci
40	907	43.1	225	20 AAW98186	Chlamydia psittaci
41	842.5	40.1	222	20 AAW98183	Chlamydia psittaci
42	825.5	39.3	215	20 AAW98185	Chlamydia psittaci
43	767	36.5	277	21 AAY82393	C. trachomatis MOM
44	519	24.7	100	20 AAW95307	Chlamydia major o
45	489	23.3	100	20 AAW95295	Chlamydia major o

ALIGNMENTS

RESULT 1
RAY35319
ID AAY35319 standard; Protein; 391 AA.
XX AC AAY35319;
XX DT 13-SEP-1999 (first entry)
XX DE Chlamydia pneumoniae transmembrane protein sequence.
XX KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
XX KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
XX KW vaccine; neutralising epitope.
XX OS Chlamydia pneumoniae.
XX PN WO9927105-A2.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-IB01890.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX (GEST) GENSET.
XX Griffiths R;
XX WPI; 1999-357842/30.
XX Genome sequence of Chlamydia pneumoniae
XX PS Page 1130-1131; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX Sequence 391 AA;

Query Match 91.5%; Score 1925; DB 20; Length 391;
Best Local Similarity 98.9%; Pred. No. 2.3e-182;
Matches 362; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 26 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 85
QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVDFCTLGASNG 121
DB 86 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVDFCTLGASNG 145
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFWSVGARGALWECGC 181
DB 146 YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFWSVGARGALWECGC 205
QY 182 ATPGAFOYAOQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 241
DB 206 ATPGAFOYAOQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 265
QY 242 NYHEWQVGSLSYRLNSLPYIGVQWSRATEDADNIRTAQPKLPTAVLNLTAWNPSSLGN 301
DB 266 NYHEWQVGSLSYRLNSLPYIGVQWSRATEDADNIRTAQPKLPTAVLNLTAWNPSSLGN 325
QY 302 ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAHV 361
DB 326 ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAHV 385
QY 362 SQGFRF 367
DB 386 SQGFRF 391

RESULT 2
AAY56771
ID AAY56771 standard; Protein; 343 AA.
XX AAY56771;
AC AAY56771;
XX 22-FEB-2000 (first entry)
XX C. trachomatis serovar HuPu MOMP sequence.
XX Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
XX cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
XX Chlamydia trachomatis.
XX WO9951745-A2.
XX 14-OCT-1999.
XX 07-APR-1999; 99WO-CA00292.
XX 07-APR-1998; 98US-0055765.
XX (UYMA-) UNIV MANITOBA.

XX Bruhnam RC;
PI WPI; 1999-620205/53.
XX Non-replicating vector encoding fragments of the outer membrane protein
XX of Chlamydia, useful in vaccines and as immunogen
XX Disclosure; Fig 10 A-F; 52pp; English.
XX The invention provides a non-replicating vector that comprises, linked
XX to a promoter, a nucleotide sequence that encodes a region containing at
XX least one of the conserved domains 2, 3 and 5 of a major outer membrane
XX protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
XX vaccines to generate a protective immune response (mainly cellular)
XX against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
XX in standard immunoassays. Immunization with the vector induces a broad
XX spectrum of immune responses, including Th1-like CD4 responses and
XX mucosal immunity, providing significant protection against subsequent
XX challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
XX of serovars of C. trachomatis.
SQ Sequence 343 AA;

Query Match 78.3%; Score 1647; DB 20; Length 343;
Best Local Similarity 86.3%; Pred. No. 7.2e-155;
Matches 316; Conservative 2; Mismatches 2; Indels 46; Gaps 1;
QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 24 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVDFCTLGASNG 121
DB 84 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVDFCTLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFWSVGARGALWECGC 181
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYDTFSFWSVGARGALWECGC 193
QY 182 ATPGAFOYAOQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 241
DB 194 ATPGAFOYAOQSKPKVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKSAI 217
QY 242 NYHEWQVGSLSYRLNSLPYIGVQWSRATEDADNIRTAQPKLPTAVLNLTAWNPSSLGN 301
DB 218 NYHEWQVGSLSYRLNSLPYIGVQWSRATEDADNIRTAQPKLPTAVLNLTAWNPSSLGN 277
QY 302 ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAHV 361
DB 278 ATALSTTDSFDFMQIVSCQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAHV 337
QY 362 SQGFRF 367
DB 338 NSGFRF 343

RESULT 3
AAW98188
ID AAW98188 standard; Protein; 389 AA.
XX AAW98188;
XX 05-JUL-1999 (first entry)
XX Chlamydia psittaci major outer membrane protein.
XX Major outer membrane protein; MOMP; psittacosis; infection;
XX vaccine; genetic immunisation.
XX Chlamydia psittaci.

```

PN WO9910005-A1.
XX 04-MAR-1999.
XX 28-AUG-1998; 98WO-US17943.
XX 28-AUG-1997; 97US-0057147.
XX (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
XX Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
PI WPI; 1999-254214/21.
XX DR N-PSDB; AAX25047.
XX A new vaccine for Chlamydia psittaci infections
XX Disclosure; Page 60-61; 72pp; English.
XX
XX The present sequence is the major outer membrane protein (MOMP)
CC of Chlamydia psittaci strain B577. A claimed MOMP polypeptide (see
CC AA98184) comprises regions V03 and V04 of B577 MOMP, i.e. it lacks
CC regions V01 and V02. A claimed vaccine composition includes MOMP
CC polypeptide lacking V01 and V02, optionally fused to a maltose
CC binding protein. Also claimed are an isolated nucleic acid
CC encoding the polypeptide, a vector, and a method of preventing C.
CC psittaci infection by administering the vaccine containing the
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking
CC regions V01 and V02 are useful for genetic vaccination. The
CC vaccines are used to prevent C. psittaci infection, especially in
XX birds.
XX
XX Sequence 389 AA;
XX
XX Query Match 73.0%; Score 1534.5; DB 20; Length 389;
XX Best Local Similarity 75.7%; Pred. No. 1.3e-143;
XX Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;
XX
QY 2 LPVGNPSPSLIDGTIWEGAGDPCDPCATWCDATSLRAGFYGYDFRILKVDAPKTF 61
DB 23 LPVGNPAEPLIDGTWEGASGDCPCDPCATWCDATSLRAGFYGYDFRILKVDVNTKI 82
QY 62 S-MGAKP--TGSAA-ANYTTAVDRPNAYNKHLDHAEFTNAGFTALNIWDRFVCTL 120
DB 83 TGMGAVPTGTAANYKTPTDRPNIAVGHQDGAFTNAAFIALNIWDRFIDFCTL 142
QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLPSVLSNGVVELYTDTSFSVSGARGALWECG 180
DB 143 GYFKASSAANLVGLIGVKGSSIAADQIPNVGITGIVGFYCDTTFSVSGARGALWECG 202
QY 181 CATLGAEFYQAQSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGVATATGTSAT 240
DB 203 CATLGAEFYQAQSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGVATATGTSAT 262
QY 241 INYHEWQVGCASLSYRLNSLVPYGVQWSRATFDADNIRIAQPKLTAVNLTAWNP 300
DB 263 IKYHEWQVGLAAYRLNSLVPYGVQWSRATFDADNIRIAQPKLTAVNLTAWNP 322
QY 301 NATALSTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINERAH 360
DB 323 EATALDTSNKFADFIQIASIQINKMKSRKACGVAVGATLIDADKWSITGEARLINERAH 382
QY 361 VSGQFRF 367
DB 383 mnaqfrf 389
XX
XX RESULT 4
XX AA98189
XX ID AA98189 standard; Protein; 402 AA.
XX
XX AC AA98189;

```

```

XX 05-JUL-1999 (first entry)
XX Chlamydia psittaci major outer membrane protein.
XX Major outer membrane protein; MOMP; psittacosis; infection;
XX vaccine; genetic immunisation.
XX Chlamydia psittaci.
XX WO9910005-A1.
XX 04-MAR-1999.
XX 28-AUG-1998; 98WO-US17943.
XX 28-AUG-1997; 97US-0057147.
XX (LOU ) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
XX Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
XX WPI; 1999-254214/21.
XX DR N-PSDB; AAX25048.
XX A new vaccine for Chlamydia psittaci infections
XX Disclosure; Page 65-66; 72pp; English.
XX
XX The present sequence is the major outer membrane protein (MOMP)
CC of Chlamydia psittaci strain 6BC. Claimed MOMP polypeptides (see
CC AA98183 and AA98184) comprise regions V03 and V04 of a MOMP, i.e.
CC they lack regions V01 and V02. Claimed vaccine compositions
CC include such MOMP polypeptides, optionally fused to a maltose
CC binding protein. Also claimed are isolated nucleic acids encoding
CC the polypeptide, a vector, and a method of preventing C. psittaci
CC infection by administering the vaccine containing the MOMP
CC polypeptide. Vectors encoding MOMP polypeptides lacking regions
CC V01 and V02 are useful for genetic vaccination. The vaccines are
CC used to prevent C. psittaci infection, especially in birds.
XX
XX Sequence 402 AA;
XX
XX Query Match 67.0%; Score 1408; DB 20; Length 402;
XX Best Local Similarity 70.3%; Pred. No. 4.7e-131;
XX Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;
XX
QY 2 LPVGNPSPSLIDGTIWEGAGDPCDPCATWCDATSLRAGFYGYDFRILKVDAPKTF 61
DB 23 LPVGNPAEPLIDGTWEGASGDCPCDPCATWCDATSLRAGFYGYDFRILKVDVNTKI 82
QY 62 S-MGAKP--TGSAA-ANYTTAVDRPNAYNKHLDHAEFTNAGFTALNIWDRFVCTL 116
DB 83 SGMAATPQTGTAANTHQPEANGRPNIAYGRHMDAEIFNSAFLNIWDRFIDFCTL 142
QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANLPSVLSNGVVELYTDTSFSWSVCA 172
DB 143 GASNGYFKASSAANLVGLIGFSAASSISTDIPMLPNVIGITGVVFEYTDTSFSWSVCA 202
QY 173 RGALWECGATLGAEFYQAQSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGV 230
DB 203 RGALWECGATLGAEFYQAQSKPKVEELNVCNVSQFVNKPKGYGVAPPLPTDAGV 262
QY 231 AFATGTSATINYHEWQVGCASLSYRLNSLVPYGVQWSRATFDADNIRIAQPKLTAVNL 290
DB 263 TEATGTSATIKYHEWQVGLAAYRLNSLVPYGVQWSRATFDADNIRIAQPKLTAVNL 322
QY 291 LTAWNPSSLIGNATAL---STTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSL 347
DB 323 ITTWNPSLIGSTALPNNSGKDVSLDIQIASIQINKMKSRKACGVAVGATLIDADKWSI 382
QY 348 TAEARLINERAHVSGQFRF 367

```


Query Match 67.0%; Score 1408; DB 22; Length 525;
 Best Local Similarity 70.3%; Pred. No. 7, 1e-131;
 Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDSLLIDGTWEGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
 Db 146 lpgvnpaepslldgtwewegagdpdpcatwcdaisiragvygdyvdrivkdvntkf 205

QY 62 S-MGAKPT---GSAA--ANTTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116
 Db 206 sgmaatptcaginasntnqpeangrpnlaygrhmqdaefnsaafialniwdrfcl 265

QY 117 GASNGYIRGNSTAFNLVGLFGVKG-----TTVNANELPNVSLNGVVELYDTSFSSVGA 172
 Db 266 gasngyfkassaafnlvlgfssaassistdplmqlpnvgitqgvvfytdtsfswsvga 325

QY 173 RGALWECGCATGAEFYAQSKPKVEELNVICNVSFVSNKPKYKGVGA--PPLPTDAGV 230
 Db 326 rgawecgcatagaefyagsqnpkiemlnvtspaqfivhkpgrkgassnplpita 385

QY 231 ATATGTSKATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
 Db 386 teatdtktsatikyhewqvglaalsyrlnmlypvgvnwsratfdadtiriaqpkkse 445

QY 291 LTAWNPSLLGNATAL---STTDSFDFMOIVSCQINKFKSRKACGVTVGATLVADKWSL 347
 Db 446 itwnpsllgstaltalpnnsqkdvlsdvlqiasiqinkmksrkacgvavgatli 505

QY 348 TAEARLINERAHVSGQRF 367
 Db 506 tgearlineraahmaaqfrf 525

RESULT 7
 AAW98187
 ID AAW98187 standard; Protein; 402 AA.
 XX AAW98187;
 AC AAW98187;
 XX AAW98187;
 DT 05-JUL-1999 (first entry)
 XX Chlamydia psittaci major outer membrane protein.
 XX Major outer membrane protein; MOMP; psittacosis; infection;
 KW vaccine; genetic immunisation.
 XX Chlamydia psittaci.
 OS WO9910005-A1.
 XX WO9910005-A1.
 PN 04-MAR-1999.
 PD 28-AUG-1998; 98WO-US17943.
 XX 28-AUG-1997; 97US-0057147.
 XX (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
 PA Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
 PI WPI; 1999-254214/21.
 DR N-PSDB; AAX25046.
 XX A new vaccine for Chlamydia psittaci infections
 PT Disclosure; Page 56-57; 72pp; English.
 PS The present sequence is the major outer membrane protein (MOMP)
 CC of Chlamydia psittaci strain LSUWCK, a cockatiel isolate (the
 CC MOMP gene sequence of this isolate is identical to that of C.
 CC psittaci Avian Type C). A claimed MOMP polypeptide (see AAW98183)
 CC comprises regions VD3 and VD4 of LSUWCK MOMP, i.e. it lacks
 CC regions VD1 and VD2. A claimed vaccine composition includes MOMP

CC polypeptide lacking VD1 and VD2, optionally fused to a maltose
 CC binding protein. Also claimed are an isolated nucleic acid
 CC encoding the polypeptide, a vector, and a method of preventing C.
 CC psittaci infection by administering the vaccine containing the
 CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking
 CC regions VD1 and VD2 are useful for genetic vaccination. The
 CC vaccines are used to prevent C. psittaci infection, especially in
 CC birds.

XX Sequence 402 AA;
 SQ

Query Match 66.9%; Score 1407; DB 20; Length 402;
 Best Local Similarity 70.0%; Pred. No. 6e-131;
 Matches 266; Conservative 39; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDSLLIDGTWEGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
 Db 23 lpgvnpaepslldgtwewegagdpdpcatwcdaisiragvygdyvdrivkdvntkf 82

QY 62 S-MGAKPT---TGSAA--ANTTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116
 Db 83 sgmaatptcaginasntnqpeangrpnlaygrhmqdaefnsaafialniwdrfcl 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG-----TTVNANELPNVSLNGVVELYDTSFSSVGA 172
 Db 143 gasngyfkassaafnlvlgfssaassistdplmqlpnvgitqgvvfytdtsfswsvga 202

QY 173 RGALWECGCATGAEFYAQSKPKVEELNVICNVSFVSNKPKYKGVGA--PPLPTDAGV 230
 Db 203 rgawecgcatagaefyagsqnpkiemlnvtspaqfivhkpgrkgassnplpita 262

QY 231 ATATGTSKATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
 Db 263 teatdtktsatikyhewqvglaalsyrlnmlypvgvnwsratfdadtiriaqpkkse 322

QY 291 LTAWNPSLLGNATAL---STTDSFDFMOIVSCQINKFKSRKACGVTVGATLVADKWSL 347
 Db 323 itwnpsllgstaltalpnnsqkdvlsdvlqiasiqinkmksrkacgvavgatli 382

QY 348 TAEARLINERAHVSGQRF 367
 Db 383 tgearlineraahmaaqfrf 402

RESULT 8
 AAY56767
 ID AAY56767 standard; Protein; 387 AA.
 XX AAY56767;
 AC AAY56767;
 DT 22-FEB-2000 (first entry)
 XX C. trachomatis serovar MoPn MOMP sequence.
 DE Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 XX cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
 KW Chlamydia trachomatis.
 OS WO9951745-A2.
 PN 14-OCT-1999.
 XX 07-APR-1999; 99WO-CA00292.
 PF 07-APR-1998; 98US-0055765.
 XX (UYMA-) UNIV MANITOBA.
 PA Bruhnam RC;
 PI WPI; 1999-620205/53.
 DR

XX Non-replicating vector encoding fragments of the outer membrane protein
 PT of Chlamydia, useful in vaccines and as immunogen
 XX
 XX
 PS Disclosure; Fig 10 A-F; 52pp; English.
 XX
 CC The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX
 SQ Sequence 387 AA;

Query Match 64.5%; Score 1356.5; DB 20; Length 387;
 Best Local Similarity 66.4%; Pred. No. 5.8e-126;
 Matches 245; Conservative 51; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPDLLIDGTIWEAGAGDCPCATWCDAISLRAGFYGYVDFDRILKVDAPKTF 61
 DB 23 lpgvnpaepsimldgilwegfggpcpcttwdcdaaisrlgyygdvfrvdktdvknkf 82
 QY 62 SNGAKPTGSAANYTAA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVCTIGA 118
 DB 83 emgaaptgd--adlttaptasrenpaygkhmqdaemftnaaymainiwdrfdvctiga 140
 QY 119 SNGYIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYTDTSFWSVGARGALWE 178
 DB 141 tsgylkgsaafnlvgldetavaadipnvsisqavveltyctatfawsgvargaalwe 200
 QY 179 CGCATLGAEFQYAOQKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKS 238
 DB 201 cgcatalgasfyagqskpkveelnvlncaaeftinkpkgyvqgefplinkagtvsatdckd 260
 QY 239 ATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQKLPFAVLMLTAWNPSL 298
 DB 261 asidihewgaslalsyrlnmftpyigvkwrsasfadtiriaqpkletsilkmttwnpti 320
 QY 299 LGNATALSTDSFDMQIVSCQINKKSRKACGVTGATLVADADKWSLTAEARLINERA 358
 DB 321 ssggidvdt--kitdtqlvslqlnkmksrksclgaigtivdvadkyavtvetrilidera 378
 QY 359 AHVSGQFRF 367
 DB 379 ahvnaqfrf 387

RESULT 9
 AAY56768
 ID AAY56768 standard; Protein; 404 AA.

AC AAY56768;

XX 22-FEB-2000 (first entry)

DE C. trachomatis serovar SFPD MOMP sequence.

KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.

OS Chlamydia trachomatis.

PN W09951745-A2.

XX 14-OCT-1999.

XX

PF 07-APR-1999; 99WO-CA00292.
 PR 07-APR-1998; 98US-0055765.
 XX (UYMA-) UNIV MANITOBA.
 XX Bruhnam RC;
 PI WPI; 1999-620205/53.
 DR Non-replicating vector encoding fragments of the outer membrane protein
 PT of Chlamydia, useful in vaccines and as immunogen
 XX
 PS Disclosure; Fig 10 A-F; 52pp; English.
 CC The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX
 SQ Sequence 404 AA;

Query Match 64.1%; Score 1349; DB 20; Length 404;
 Best Local Similarity 64.8%; Pred. No. 3.4e-125;
 Matches 248; Conservative 51; Mismatches 66; Indels 18; Gaps 4;

QY 2 LPVGNPSPDLLIDGTIWEAGAGDCPCATWCDAISLRAGFYGYVDFDRILKVDAPKTF 61
 DB 23 lpgvnpaepsimldgilwegfggpcpcttwdcdaaisrlgyygdvfrvdktdvknkf 82
 QY 62 SNGAKPTGSG---AAANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVCTIGA 118
 DB 83 emgpypttdtdadaadittstprenpaygkhmqdaemftnaaymainiwdrfdvctiga 142
 QY 119 SNGYIRGNSTAFNLVGLFGVKG-----TTVNANELPNVSLNSGVVVELYTDTSFWSV 170
 DB 143 tsgylkgsaafnlvgldetavaadipnvsisqavveltyctatfawsgvargaalwe 201
 QY 171 GARGALWEGCGCATLGAEFQYAOQKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGV 230
 DB 202 garaalwegcgcatlgaefqyagqskpkveelnvlncaaeftinkpkgyvkefplaltag 261
 QY 231 ATATGCTKATINHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQKLPFAVLN 290
 DB 262 dsatdkdasidynhewgaslalsyrlnmftpyigvkwrsasfadtiriaqpklaealid 321
 QY 291 LTAWNPSLLGNAT-----ALSTDSFDMQIVSCQINKKSRKACGVTGATLVADADK 344
 DB 322 vttwnptiagagtiadgtgaaatangladtlqivslqlnkmksrksclgaigtivdvadk 381
 QY 345 WSLTAEARLINERAHVSGQFRF 367
 DB 382 yavtvetrilideraahvnaqfrf 404

RESULT 10
 AAY76366
 ID AAY76366 standard; protein; 376 AA.

XX AAY76366;

XX 03-DEC-1998 (first entry)

XX C. trachomatis JM109 pCT33-H major outer membrane protein.

XX

KW Major outer membrane protein; antibody; antigen; diagnosis; disease.
 XX Chlamydia trachomatis.
 OS JPI0234395-A.
 XX
 PN 08-SEP-1998.
 XX
 PD 25-FEB-1997; 97JP-0040780.
 XX
 PF 25-FEB-1997; 97JP-0040780.
 XX
 PR (ELED) DENKI KAGAKU KOGYO KK.
 XX
 PA WPI; 1998-535045/46.
 XX
 DR Preparation of major outer membrane protein of Chlamydia trachomatis
 XX - by cloning and recombinant expression of the gene, for use as a
 PT diagnostic of Chlamydia infections
 XX
 PS Claim 2; Page 15-16; 19pp; Japanese.
 XX
 CC AAM76362-W76366 are major outer membrane proteins isolated from
 CC Chlamydia trachomatis which are used for the measurement of an antibody
 CC against Chlamydia trachomatis, by using it as an antigen in the form of
 CC a reagent. The method can give a diagnostic method for Chlamydia
 CC trachomatis-infected diseases with high specificity.
 XX
 SQ Sequence 376 AA;

Query Match 64.0%; Score 1345.5; DB 19; Length 376;
 Best Local Similarity 65.2%; Pred. No. 6.8e-125;
 Matches 245; Conservative 51; Mismatches 71; Indels 9; Gaps 3;
 Qy 1 MLPVGNPSDLLIDGTWEGAGDPCDPCATWCDATSLRAGFYGDYVDFRLKVDAPKT 60
 Db 1 MLPVGNPAEPSIMIDGILWEGFGGDPDPCATWCDATSLMRVYGGDFRVLKDVNKE 60
 Qy 61 FSMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEFTNAGFTALNWDREDFVCTL 116
 Db 61 fmggaaptndaaadlqndpknvarpnpaygkqhmqdaemftnaaymalniwdrfdvctl 120
 Qy 117 GASNGYIRGNSTAFNLVGLFGVKGKTTVNANE---LPNVSLNGVVELYDTFSFWSVGAR 173
 Db 121 gattgylkgnasnlvlgftkssdfntaklvpnlalnrvvlydttdttfawsvgar 180
 Qy 174 GALWECGCATLCAEFOYQASPKVBEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATA 233
 Db 181 aalwecgcatalgasfygaqskpkveelnvlnaseftinkpkgygaefpdlitagtea 240
 Qy 234 GTGKATINYHEWQVGSLSRLNLSVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAA 293
 Db 241 tktkdsidyhewqaslsrlnmftpygkwsrvsfadadtiriaqpklaeallvtt 300
 Qy 294 WNPSSLGNAT--ALSTTDSFDMQIVSCQINKFKSRKACGVTVGCATLVADKWSLTAE 351
 Db 301 lnptiagktvvasgsdndlqtmqivslqlnkmksrksclgiavgttvdadkyavtvet 360
 Qy 352 RLINERAAHVSGQFRF 367
 Db 361 rlideraahvnaqfrf 376

RESULT 11
 AAY56766
 ID AAY56766 standard; Protein; 397 AA.
 XX
 AC AAY56766;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE C. trachomatis serovar H MOMP sequence.

XX Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 KW cellular response; immunogen; Thi-like CD4 response; mucosal immunity.
 XX Chlamydia trachomatis.
 OS WO9951745-A2.
 XX
 PN 14-OCT-1999.
 XX
 PD 07-APR-1999; 99WO-CA00292.
 XX
 PF 07-APR-1998; 98US-0055765.
 XX
 PR (UYMA-) UNIV MANITOBA.
 XX
 PA Bruhnam RC;
 XX
 PI WPI; 1999-620205/53.
 XX
 DR Non-replicating vector encoding fragments of the outer membrane protein
 XX of Chlamydia, useful in vaccines and as immunogen
 PT
 PT Disclosure; Fig 10 A-F; 52pp; English.
 XX
 PS The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Thi-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX
 SQ Sequence 397 AA;

Query Match 63.6%; Score 1338.5; DB 20; Length 397;
 Best Local Similarity 64.8%; Pred. No. 3.7e-124;
 Matches 243; Conservative 52; Mismatches 71; Indels 9; Gaps 3;
 Qy 2 LPVGNPSDLLIDGTWEGAGDPCDPCATWCDATSLRAGFYGDYVDFRLKVDAPKTF 61
 Db 23 lpgvnpaepsimidgilwegfggdpdpcatwcdaismrvggygdfvdrvlktdvnkef 82
 Qy 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEFTNAGFTALNWDREDFVCTLG 117
 Db 83 qmggaaptndaaadlqndpknvarpnpaygkqhmqdaemftnaaymalniwdrfdvctlg 142
 Qy 118 ASNGYIRGNSTAFNLVGLFGVKGKTTVNANE---LPNVSLNGVVELYDTFSFWSVGARG 174
 Db 143 attgylkgnasnlvlgftkssdfntaklvpnlalnrvvlydttdttfawsvgara 202
 Qy 175 ALWECGCATLCAEFOYQASPKVBEELNVICNVQSFVNKPKGYKGVAFPLPTDAGVATAT 234
 Db 203 alwecgcatalgasfygaqskpkveelnvlnaseftinkpkgygaefpdlitagtea 262
 Qy 235 GTKATINYHEWQVGSLSRLNLSVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAA 294
 Db 263 gtkdsidyhewqaslsrlnmftpygkwsrvsfadadtiriaqpklaeallvttl 322
 Qy 295 NPSLLGNAT--ALSTTDSFDMQIVSCQINKFKSRKACGVTVGCATLVADKWSLTAEAR 352
 Db 323 nptiagktvvasgsdndlqtmqivslqlnkmksrksclgiavgttvdadkyavtvetr 382
 Qy 353 LINERAAHVSGQFRF 367
 Db 383 rlideraahvnaqfrf 397

RESULT 12
 AAY56763
 ID AAY56763 standard; Protein; 397 AA.
 XX
 AC AAY56763;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE C. trachomatis serovar L3 MOMP sequence.
 XX
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WO9951745-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-CA00292.
 XX
 PR 07-APR-1998; 98US-0055765.
 XX
 PA (UYMA-) UNIV MANITOBA.
 XX
 PI Bruhnam RC;
 XX
 DR WPI; 1999-620205/53.
 XX
 PT Non-replicating vector encoding fragments of the outer membrane protein
 PT of Chlamydia, useful in vaccines and as immunogen
 XX
 PS Disclosure; Fig 10 A-F; 52pp; English.
 XX
 CC The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX
 SQ Sequence 397 AA;

Query Match 63.6%; Score 1336.5; DB 20; Length 397;
 Best Local Similarity 64.8%; Pred. No. 5.8e-124;
 Matches 243; Conservative 50; Mismatches 73; Indels 9; Gaps 3;
 QY 2 LPVGNPSPDLLIDGTWEGAAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
 Db 23 lpvgnpaeplsimidgilwegf9gdpctctwcdaismrvygdvfrvdktdvnef 82
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFVDFCTLG 117
 Db 83 qmgaepstsdtaglsnldpttnvarpnayghmqdaemftnaaymalniwdrfvdvctlg 142
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGKTVNANE---LPNVSLNSNGVVELYTDTSFWSVGARG 174
 Db 143 attgylkgnasfalsvlgftkqstnfntaklvpntalnqavveltydttdttfawsvgara 202
 QY 175 ALWEGCATLCAEFOYAGSKPKVELNVCNVSQFSVKNPKGYGVAFPLPTDAGVATAT 234
 Db 203 alwecgcatlgasfgyaqskpkveelnvldcaesftlnkpkgygvaefplditagteat 262
 QY 235 GTSKATINHYEQVAGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAW 294
 Db 263 gtdkasiyehewqslalsylrlnmftpyigvkwsvsfadattiriaqpklaeavldvttl 322

QY 295 NPSLL--GNATALSTDSFDMQIVSCQINKEKSRKAGVTGATLVADKWSLTARAR 352
 Db 323 nptiagksgvsvasgseneladtumqivslqlnkmksrkgciavgttiadvadkyavivetr 382
 QY 353 LINERAAHVSGQFRF 367
 Db 383 llderaahvnaqfrf 397
 RESULT 13
 AAY56764
 ID AAY56764 standard; Protein; 396 AA.
 XX
 AC AAY56764;
 XX
 DT 22-FEB-2000 (first entry)
 XX
 DE C. trachomatis serovar A MOMP sequence.
 XX
 KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
 XX
 OS Chlamydia trachomatis.
 XX
 PN WO9951745-A2.
 XX
 PD 14-OCT-1999.
 XX
 PF 07-APR-1999; 99WO-CA00292.
 XX
 PR 07-APR-1998; 98US-0055765.
 XX
 PA (UYMA-) UNIV MANITOBA.
 XX
 PI Bruhnam RC;
 XX
 DR WPI; 1999-620205/53.
 XX
 PT Non-replicating vector encoding fragments of the outer membrane protein
 PT of Chlamydia, useful in vaccines and as immunogen
 XX
 PS Disclosure; Fig 10 A-F; 52pp; English.
 XX
 CC The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX
 SQ Sequence 396 AA;

Query Match 63.1%; Score 1326; DB 20; Length 396;
 Best Local Similarity 64.7%; Pred. No. 6.3e-123;
 Matches 242; Conservative 48; Mismatches 76; Indels 8; Gaps 3;

QY 2 LPVGNPSPDLLIDGTWEGAAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
 Db 23 lpvgnpaeplsimidgilwegf9gdpctctwcdaismrvygdvfrvdktdvnef 82
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFVDFCTLG 117
 Db 83 qmgaepstsdtaglsnldpttnvarpnayghmqdaemftnaaymalniwdrfvdvctlg 142
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGKTV---NANLPLNVSLNSNGVVELYTDTSFWSVGARG 174
 Db 143 attgylkgnasfalsvlgftkqstnfntaklvpntalnqavveltydttdttfawsvgara 202


```

Db      1 mlpvgnpaepsImldgIlwegfggpcpcttwcdaismrvggygdfvdrvlktdvke 60
Qy     61 FSMGAKPTGSAANY---TTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFCFL 116
Db     61 fmgaapttsdvagIqndptInvarpnpaykHmqdaemftnaaymalniwdrfadvctl 120
Qy    117 GASNGYIRGNSTAFNLVGLFGVKGTVNANE---LPNVSLSNGVVVELYTDTSFSWSVGAR 173
Db    121 gattgylkgnsaasfnlvgIfgktqssfnaklipntalneavvelyInttfawsvgar 180
Qy    174 GALWECGCATLCAEFOYAQSFPKVEELNVICNVSOFSYKPKGYKGVAFPLPTDAGVATA 233
Db    181 aalwecgcaticgasfgyagskpveelnvIcnaseftInkpgyvgaeFpInitagteaa 240
Qy    234 TGTKSATINYHEWQYGASLSYRLNSLVPYIGVOWSRATFDADNIRIAOPKLPtAVLNlTA 293
Db    241 tgtkdasIdynewqaslsalsyrlmmftpyIgvkwsrsvfdadtIrIagpkIaealldvtt 300
Qy    294 WNPSLLGNATALS--TTDSFSDFMQIVSCQINKFKSRKACGVTVGATLVADADKWSLTAEa 351
Db    301 InrttagkvsvsagtDneladtmqIvslqInkmksrkgIavgttIvdadkyavtvea 360
Qy    352 RLINERAAHVSCQFRF 367
Db    361 rliDeraahvnaqfrf 376

```

Search completed: February 7, 2002, 21:34:40
Job time: 20759 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:36:24 ; Search time 85.91 seconds
(without alignments)
103.204 Million cell updates/sec

Title: US-09-391-606-15

Perfect score: 2103

Sequence: 1 MLPVGNFSPDLSLLIDTWE.....QKLISEDLNSAVDHHHHH 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA.*

1: /cgn2_6/ptodata/2/1aa/5A_COMB.pap.*

2: /cgn2_6/ptodata/2/1aa/5B_COMB.pap.*

3: /cgn2_6/ptodata/2/1aa/6A_COMB.pap.*

4: /cgn2_6/ptodata/2/1aa/6B_COMB.pap.*

5: /cgn2_6/ptodata/2/1aa/PCTUS_COMB.pap.*

6: /cgn2_6/ptodata/2/1aa/backfiles1.pap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	142.5	6.8	1154	US-08-974-549A-611	Sequence 611, Appl
2	136.5	6.5	61	US-08-374-560-1	Sequence 1, Appl
3	125	5.9	42	PCT-US93-08739-3	Sequence 3, Appl
4	124	5.9	25	US-08-446-692-45	Sequence 45, Appl
5	124	5.9	25	US-08-488-351A-45	Sequence 45, Appl
6	124	5.9	25	PCT-US93-08739-1	Sequence 1, Appl
7	124	5.9	25	PCT-US95-13841-17	Sequence 17, Appl
8	124	5.9	27	US-09-411-329C-22	Sequence 22, Appl
9	124	5.9	27	US-09-411-329C-26	Sequence 26, Appl
10	124	5.9	37	US-08-446-692-24	Sequence 24, Appl
11	124	5.9	37	US-08-488-351A-24	Sequence 24, Appl
12	124	5.9	439	US-09-433-428D-57	Sequence 57, Appl
13	121	5.8	368	US-09-433-428D-58	Sequence 58, Appl
14	119	5.7	31	US-08-519-385B-1	Sequence 1, Appl
15	115	5.5	25	US-09-100-409A-51	Sequence 51, Appl
16	106.5	5.1	553	US-08-661-052-16	Sequence 16, Appl
17	106.5	5.1	553	US-09-188-082-16	Sequence 16, Appl
18	101.5	4.8	919	US-08-788-674-4	Sequence 4, Appl
19	101	4.8	289	US-09-184-658-63	Sequence 63, Appl
20	100	4.8	21	US-09-071-710-41	Sequence 41, Appl
21	100	4.8	21	US-08-912-276-25	Sequence 25, Appl
22	100	4.8	21	US-09-153-804-17	Sequence 17, Appl
23	100	4.8	21	US-08-944-483-76	Sequence 76, Appl
24	100	4.8	21	US-09-525-397-41	Sequence 41, Appl
25	99	4.7	57	US-08-958-201-18	Sequence 18, Appl
26	98	4.7	23	US-08-350-260A-42	Sequence 42, Appl
27	98	4.7	57	US-08-652-816A-41	Sequence 41, Appl

28 98 4.7 301 2 US-08-661-052-14 Sequence 14, Appl
29 98 4.7 301 4 US-09-188-082-14 Sequence 14, Appl
30 97.5 4.6 457 3 US-09-142-759-1 Sequence 1, Appl
31 90.5 4.3 784 4 US-09-371-913A-7 Sequence 7, Appl
32 89 4.2 1199 4 US-09-208-742-2 Sequence 2, Appl
33 89 4.2 1199 4 US-09-332-295-4 Sequence 4, Appl
34 87.5 4.2 505 3 US-08-627-907A-2 Sequence 2, Appl
35 87.5 4.2 834 1 US-08-471-033-21 Sequence 21, Appl
36 87.5 4.2 834 2 US-08-471-044-21 Sequence 21, Appl
37 87.5 4.2 834 2 US-08-463-483A-21 Sequence 21, Appl
38 87.5 4.2 834 2 US-08-471-046A-21 Sequence 21, Appl
39 87.5 4.2 834 2 US-08-470-566B-21 Sequence 21, Appl
40 87.5 4.2 834 2 US-08-469-334-21 Sequence 21, Appl
41 87.5 4.2 834 3 US-09-300-529-21 Sequence 21, Appl
42 86 4.1 512 4 US-08-856-253-6 Sequence 6, Appl
43 86 4.1 1183 2 US-08-447-031A-2 Sequence 2, Appl
44 85 4.0 621 3 US-08-262-220-8 Sequence 8, Appl
45 85 4.0 621 4 US-08-750-494-8 Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-974-549A-611
; Sequence 611, Application US/08974549A
; Patent No. 6166178
; GENERAL INFORMATION:
; APPLICANT: Cech, Thomas R.
; APPLICANT: Lingner, Joachim
; APPLICANT: Nakamura, Toru
; APPLICANT: Chapman, Karen B.
; APPLICANT: Morin, Gregg B.
; APPLICANT: Harley, Calvin B.
; APPLICANT: Andrews, William H.
; TITLE OF INVENTION: Human Telomerase Catalytic Subunit
; NUMBER OF SEQUENCES: 727
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend and Crew LLP
; STREET: Two Embarcadero Center, Eighth Floor
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94111-3834

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/974,549A
FILING DATE: 19-NOV-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/724,643
FILING DATE: 01-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/844,419
FILING DATE: 18-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/846,017
FILING DATE: 25-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/851,843
FILING DATE: 06-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/854,050
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/911,312
FILING DATE: 14-AUG-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/912,951

;; FILING DATE: 14-AUG-1997
;; PRIOR APPLICATION DATA: US 08/915,503
;; FILING DATE: 14-AUG-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US97/17618
;; FILING DATE: 01-OCT-1997
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: WO PCT/US97/17885
;; FILING DATE: 01-OCT-1997
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Apple, Randolph Ted
;; REGISTRATION NUMBER: 36,429
;; REFERENCE/DOCKET NUMBER: 015389-002610US
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (415) 576-0200
;; TELEFAX: (415) 576-0300
;; INFORMATION FOR SEQ ID NO: 611:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 1154 amino acids
;; TYPE: amino acid
;; STRANDEDNESS:
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; FEATURE:
;; NAME/KEY: Protein
;; LOCATION: 1..1154
;; OTHER INFORMATION: /note= "fusion protein composed of hprt
;; protein sequence, vector sequences, the
;; OTHER INFORMATION: Myc epitope and His6 tag"
US-08-974-549A-611

Query Match 6.8%; Score 142.5; DB 4; Length 1154;

Best Local Similarity 26.18; Pred. No. 9.3e-06;

Matches 54; Conservative 19; Mismatches 37; Indels 97; Gaps 7;

QY 200 LNVICNVSO--FSVKNPK-----GYKGVAFPLPTDAGVATATGKTSATINVHEWQVGAS 251

DB 1033 LVAISDTASLCYSILKAKNAGSLGAKGAGPLPSEA-----VQWLCHQA 1077

QY 252 LSYRLN-----SLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATALST 307

DB 1078 FLKLTRHRTVYVPLLG-----SLRTAQTL----- 1103

QY 308 TDSFDFMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAERLINERAHHVSGQFRF 367

DB 1104 -----SRKLPGTTTLAL-----EAAANPALPSDF 1127

QY 368 RYQAYVEQKLISEEDLNSAVDHHHHH 394

DB 1128 KTLIDLEQKLISEEDLNSAVDHHHHH 1154

RESULT 2

US-08-374-560-1

; Sequence 1, Application US/08374560

; Patent No. 5882645

; GENERAL INFORMATION:

; APPLICANT: TOTH, Istvan

; APPLICANT: GIBBONS, William Anthony

; TITLE OF INVENTION: PEPTIDE COMPOUNDS

; NUMBER OF SEQUENCES: 39

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen

; STREET: 1180 Avenue of the Americas

; CITY: New York

; STATE: NY

; COUNTRY: US

; ZIP: 10036-8403

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/374,560

FILING DATE: 13-MAR-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: GB 9215780.9

FILING DATE: 24-JUL-1992

ATTORNEY/AGENT INFORMATION:

NAME: Meilman, Edward A.

REGISTRATION NUMBER: 24,735

REFERENCE/DOCKET NUMBER: P/365-302

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 382-0700

TELEFAX: (212) 382-0888

TELEX: 236925

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 61 amino acids

TYPE: amino acid

STRANDEDNESS:

TOPOLOGY: linear

MOLECULE TYPE: peptide

FRAGMENT TYPE: internal

US-08-374-560-1

Query Match 6.5%; Score 136.5; DB 2; Length 61;

Best Local Similarity 47.5%; Pred. No. 3.4e-07;

Matches 29; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

QY 269 RATPDADNIRIAQPKLPTAVLNLTAWNPSSLGNATALSTTD-SFSDFMQIVSCQINKFKS 327

DB 1 RASFDADIRIAQPKSAETIFDVTLTNTIAGAGDVKTSAGQLGDTHTQIVSLQLNKMK 60

QY 328 R 328

DB 61 R 61

RESULT 3

PCT-US93-08739-3

; Sequence 3, Application PC/TUS9308739

; GENERAL INFORMATION:

; APPLICANT: The Government of the United States of America

; APPLICANT: as represented

; APPLICANT: by the Secretary of the Department of Health and Human Services

; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR

; TITLE OF INVENTION: CHLAMYDIA

; NUMBER OF SEQUENCES: 3

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/08739

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 07/947,671 US

; FILING DATE: 18 SEP 93

; INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 42 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHEITICAL: NO

ANTI-SENSE: NO

FRAGMENT TYPE: N-terminal
PCT-US93-08739-3

Query Match 5.98; Score 125; DB 5; Length 42;
Best Local Similarity 61.98; Pred. No. 3.1e-06;
Matches 26; Conservative 4; Mismatches 2; Indels 10; Gaps 2;

Qy 103 ALNIWDRDVFCTLGASNGYIRGNSAFNLVGLFCVKGTTVN 144
|||||
Db 1 ALNIWDRDVFCTLGATGTYLKGN-----FDV--TTLN 32

RESULT 4

US-08-446-692-45
Sequence 45, Application US/08446692
Patent No. 5759551

GENERAL INFORMATION:
APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Y1

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs

TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines

NUMBER OF SEQUENCES: 114

CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-446-692-45

Query Match 5.98; Score 124; DB 1; Length 25;
Best Local Similarity 84.08; Pred. No. 1.7e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRDVFCTLGASNGYIRGNS 127
|||||

Db 1 ALNIWDRDVFCTLGATGTYLKGN 25

RESULT 5

US-08-488-351A-45

Sequence 45, Application US/08488351A

Patent No. 5843446

GENERAL INFORMATION:

APPLICANT: Ladd, Anna

APPLICANT: Wang, Chang Y1

APPLICANT: Zamb, Timothy

TITLE OF INVENTION: Immunogenic LHRH peptide constructs
TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:

ADDRESSEE: Maria C.H. Lin

STREET: 345 Park Avenue

CITY: New York

STATE: NY

COUNTRY: US

ZIP: 10154-0053

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/488,351A

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/446,692

FILING DATE: 7-JUN-1995

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/229,275

FILING DATE: 14-APR-1994

CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/057,166

FILING DATE: 27-APR-1992

CLASSIFICATION: 424

ATTORNEY/AGENT INFORMATION:

NAME: Maria C.H. Lin

REGISTRATION NUMBER: 29,323

REFERENCE/DOCKET NUMBER: 1151-4146 US2

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212)415-8745

TELEFAX: (516)751-6849

INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:

LENGTH: 25 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-08-488-351A-45

Query Match 5.98; Score 124; DB 2; Length 25;
Best Local Similarity 84.08; Pred. No. 1.7e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRDVFCTLGASNGYIRGNS 127
|||||

Db 1 ALNIWDRDVFCTLGATGTYLKGN 25

RESULT 6

PCT-US93-08739-1

Sequence 1, Application PC/TUS9308739

GENERAL INFORMATION:

APPLICANT: The Government of the United States of America

APPLICANT: as represented

APPLICANT: by the Secretary of the Department of Health and Human Services

TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR

TITLE OF INVENTION: CHLAMYDIA

TITLE OF INVENTION: TRACHOMATIS

NUMBER OF SEQUENCES: 3

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: PCT/US93/08739
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,671 US
FILING DATE: 18 SEP 93
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
PCT-US93-08739-1

Query Match 5.9%; Score 124; DB 5; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.7e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 7

Sequence 17, Application PC/TUS9513841
GENERAL INFORMATION:
APPLICANT: United Biomedical Inc; Walfield, Alan M.;
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Synthetic IgE Membrane Anchor
TITLE OF INVENTION: Peptide Immunogens for the Treatment of Allergy
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13841
FILING DATE: 25-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/328,519
FILING DATE: 25-OCT-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lin, Maria C.H.
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4117
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13841-17

Query Match 5.9%; Score 124; DB 5; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.7e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 8

US-09-411-329C-22
Sequence 22, Application US/09411329C
Patent No. 6261820
GENERAL INFORMATION:
APPLICANT: Boone, Thomas
APPLICANT: Li, Huimin
APPLICANT: Mann, Michael
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
FILE REFERENCE: A-596
CURRENT APPLICATION NUMBER: US/09/411,329C
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.0
SEQ ID NO 22
LENGTH: 27
TYPE: PRT
ORGANISM: Agkistrodon contortrix
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: Native pro-fibrolase of Agkistrodon contortrix
US-09-411-329C-22

Query Match 5.9%; Score 124; DB 4; Length 27;
Best Local Similarity 87.5%; Pred. No. 1.9e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 371 AYVEQKLISEEDLNSAVDHHHHH 394
Db 4 SFLEQKLISEEDLNSAVDHHHHH 27

RESULT 9

US-09-411-329C-26
Sequence 26, Application US/09411329C
Patent No. 6261820
GENERAL INFORMATION:
APPLICANT: Boone, Thomas
APPLICANT: Li, Huimin
APPLICANT: Mann, Michael
TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
FILE REFERENCE: A-596
CURRENT APPLICATION NUMBER: US/09/411,329C
CURRENT FILING DATE: 1999-10-01
NUMBER OF SEQ ID NOS: 29
SOFTWARE: PatentIn version 3.0
SEQ ID NO 26
LENGTH: 27
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Analog form of native pro-fibrolase of Agkistrodon contortrix
US-09-411-329C-26

Query Match 5.9%; Score 124; DB 4; Length 27;
Best Local Similarity 87.5%; Pred. No. 1.9e-06;
Matches 21; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 371 AYVEQKLISEEDLNSAVDHHHHH 394
Db 4 SFLEQKLISEEDLNSAVDHHHHH 27

```

Query Match      5.9%; Score 124; DB 1; Length 37;
Best Local Similarity 84.0%; Pred. No. 3.2e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY   103 ALNIWDRFDVCTGLGASNGYIRGNS 127
       |||||
Db    1 ALNIWDRFDVCTGLGATTCGYLKGNs 25
       |||||

RESULT 11
US-08-488-351A-24
; Sequence 24, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Laod, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk

```

```

Query Match      5.9%; Score 124; DB 2; Length 37;
Best Local Similarity 84.0%; Pred. No. 3.2e-06;
Matches 21; Conservative 3; Mismatches 1; Indels

QY 103 ALNIWRFDVFCITGLASNGYIRGNS 127
      |||||
Db 1 ALNIWRFDVFCITGATTCTYLRGNS 25
      |||||

```

```

Query Match      5.98; Score 124; DB 4; Length 439;
Best Local Similarity 57.8%; Pred. No. 0.00017;
Matches 26; Conservative 3; Mismatches 10; Indels 6; Gaps 1;

QY 356 ERAAHVSGQFRFYQAVY-----EOKLIISEDLSNAVDHHHHHH 394
      ||| : : : :
      ||| : : : :

```

STATE: WI

ATTORNEY/AGENT INFORMATION:

NAME:

NAME: _____
: _____
ATTORNEY/AGENCY: _____NAME: _____
: _____
ATTORNEY / AGENT: _____


```

;
; REGISTRATION NUMBER:
; REFERENCE/DOCKET NUMBER: 1151-4154
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-758-4800
; TELEFAX: 212-751-6849
; INFORMATION FOR SEQ ID NO: 51:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 25 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-100-409A-51

```

```

Query Match          5.5%; Score 115; DB 3; Length 25;
Best Local Similarity 80.0%; Pred. No. 1.5e-05;
Matches 20; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
    |||||
DB 1 ALNIWDRFDVFTLGATSGYLKGN 25

```

Search completed: February 7, 2002, 21:36:25
Job time: 20494 sec


```

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
; IMMEDIATE SOURCE:
; CLONE: APC
US-08-452-654-7

```

Query Match 6.1%; Score 138; DB 1; Length 2842;
Best Local Similarity 20.0%; Pred. NO. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY	11	DETERTPPADLSA-OGLEASAANKSAEQAORIAGAEAKPKE-----SKTDSV-----	55
Db	1658	DLTIESPPNELAAGEGVRGAOSGEFERDRTIPTEGRSTDBAQOGKKTSSVTIPELDONKA	1717
QY	56	ERWSILRSVAVNALM-----SLADKLGTASSNSSSTSRSDADVSTTTATATPP	103
Db	1718	EEGDIIAECINSAMPKGSKHKPFVRYKKIWDVQOQASASSAPNKNQLDGKKKKTSPVKP	1777
QY	104	PPSDDYKTKQAOTAYDT-----IFT-----STSLADIQAALVSLQDAV----	1411
Db	1778	IPONTEYTRVRKRNADSKNNLNAERVFSDNKDSKQNLKNNSKDFNOKLPNEDVRGSEF	1837
QY	142	-----FNIKDT--AATDEETATAAEWETKNADAIKVGAGIITELAKAYASDQAIILDSL	191
Db	1838	AFDSPHHYPIEGTPYCFSRNDSLSSLOFDDDDVLSREKAEL-----KKAKENK---ESE	1890
QY	192	GKLTSEDLLQOTALLOSVAANNKKAELLKEM-----QDNVPVPGKTPATAQSLVDOTATAT	247
Db	1891	AKVTS-----HTELTSNQOQSANKTOAIKOPINRGQPKPILQ-KOSTPPOSSKIDPRGAA	1945
QY	248	QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSKAKAIAATAKTQIAEAAQKKFPDSP	307
Db	1946	TDEKLQN-----FAIENTPVCFSHNSLSUSDID-----	1975
QY	308	ILQEAQBQWIOAEKDLKNIKPADGSDVDPNGTIVGSGKQ-----QGSSIGSIR	355
Db	1976	--QENNN-----KENEPIKETEPDQSGEFSKPOAGSYAPKSPHVEDTPVCFSRNSSLSSLS	2030
QY	356	VSMLLDDAENETASILMSGFROWIHFMTENPDSQAOQELAAQARAACAAGDDSAARAAL	415
Db	2031	I-----DSEDDLLQECISS-----AMPKKKKPSRLUKGDNKHSRPNMGWILGED-LFTLDL	2079
QY	416	ADAQKA-LEAALG-----KAGQO--QGILNALQITASAAYVS	449
Db	2080	KDIORPDSBHGILSPDSENFEDKWAIOEGANSIVSSLUHAAAAACILS	2124

RESULT 11
 US-07-741-940-2
 ; Sequence 2, Application US/07741940
 ; Patent No. 5352775
 ; GENERAL INFORMATION:
 ; APPLICANT: ALBERTSEN, HANS
 ; APPLICANT: ANAND, RAKESH
 ; APPLICANT: CARLSON, MARY
 ; APPLICANT: GRODEN, JOANNA
 ; APPLICANT: HEDGE, PHILIP J.
 ; APPLICANT: JOSLYN, KENNETH
 ; APPLICANT: KINZLER, KENNETH
 ; APPLICANT: MARKHAM, ALEXANDER F.
 ; APPLICANT: NAKAMURA, YUSUKE
 ; APPLICANT: THLIVERIS, ANDREW
 ; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 ; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 ; NUMBER OF SEQUENCES: 94
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Banner, Birch, McKie & Beckett
 ; STREET: 1001 G Street, NW

CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:

```

Query Match      6.1% Score 138; DB 1; Length 2843;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY      11 DEBTEPTPADLSA-OGLEASAAKSAEORITAGAEAKPKE-----SKTDSV----- 55
Db      1659 DLITESPNELAAGEVGRGNAQSECFEKRDTIPIEGASTDEAQGGKTSVITPELDDNKA 1718
QY      56 ERWSILRSVNNALM-----SLADKLGIASSNSSTSRSDVDSTTTATAPPP 103
Db      1719 EEGDILAEICINSAMPKSHKPRVKKIMDQVOQASASSAPNKNQLDGKKKKPTSPVKP 1778
QY      104 PPTSDDYKTOAQATYDT-----IFT-----STSLADIQAALVSLQDAV--- 141
Db      1779 IPQNTETRTVRKNKADSKNNLNAERVFSDNKDKSKQNLKNKSFNDKLPNNEDVRGGSF 1838
QY      142 -----TNIKDT--AATDEETAAIAEWETKNADATKVGAQITELAKYASDNQAILDLSL 191
Db      1839 AFDSPHHYTTIEGTPYCFSRNDSLSLSDDDDDVLSREKAEL---RKAKENK---ESE 1891
QY      192 GKLTSTFDLLQTALQSVANNKAAELLEM---QDNPNVPGKTPAIAQSLVDQTDATAT 247
Db      1892 AKVTS---HTELTSNQOSANKTOAIKQPINRGOPKILQ-KOSTFPQSSKIDPRGAA 1946
QY      248 QIEKDGNAGIDAVFAGONASGAVENAKNSNISDSAKAIAATAKTOIAEAKQKFPDSP 307
Db      1947 TDEKLQN-----FAIENTPVCFSHNSLSLSLSDID----- 1976
QY      308 ILQAEQMWIOAEKDLKNIKPADGSDVPNPGTTPVGGSKQ-----QGSSIGSIR 355
Db      1977 --QENNN--KENEPKETETPPDQSEPKPQASGYAPKSFHVEDTPVCFSRNSLSLSLS 2031
QY      356 VSMLLDDAEANETASILMSGFGOMTHMNTENPNPDSQAQOELAAQARAKAAGDSDSAAAL 415
Db      2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGDNEKHSRPNMGILGED-LTLDL 2080
QY      416 ADAQKA-LEAALG-----KAGQQ--QGTLNALGQIASAAVVS 449
Db      2081 KDIORPDSEHCLSPDSENFQWKAIQEGANSIVSSLHQAAAAACLS 2125

RESULT 12
US-08-289-548A-2
Sequence 2, Application US/08289548A

```

```

Patent No. 5648212
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Allegretti, LTD
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/289,548A
FILING DATE: 12-AUG-1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.46943
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-289-548A-2

Query Match 6.18; Score 138; DB 1; Length 2843;
Best Local Similarity 20.08; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55
Db 1659 DLTIESPNEUAEAGVGGAGQSGFEKRDPTIPEGSTDEAQGGKTSSTVPIELDDNKA 1718
QY 56 ERWSILRSVAVNLM-----SLADKLGITASSNSSSTSRSDVSTTATATPTTP 103
Db 1719 EEGDILAEBCINSMPKGGKSHKPRVVKIMDQVOQASASSAPNKNQLDGKKKKPTSPVKP 1778
QY 104 PPTSDDKYKTAQYADT-----IFT-----STSLADIQAALVSLQDAV--- 141
Db 1779 IPQNTYRTRVRKNWADSKNNINAEVFSNDKSKQNKLNKNSKDFNDKLPNNEDRVRGSP 1838
QY 142 -----TNIKDT--AATDEETAIAAEWETKNAIDAKVGAQITELAKYASDNQAILDSL 191
Db 1839 AFDSPHHYTPTEGPFYCFYSRNDLSLDDFDDDDVDSREKAEL---RKAKENK---ESE 1891
QY 192 GKLTSLFDLLOLQALQSIVANNKKAELKEM-----QDNPPVPEKTPATAQSLVDQTDATAT 247
Db 1892 AKVTS-----HTLSTNSQOSANKTQIAKQPIINRGQPKPILO-KOSTFPQSSKIDPDRGAA 1946
QY 248 QIEKDGNAGIDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAQKFPDSP 307

```

```

Db 1947 TDEKLQ-----FAIENTPVCFESHNSLSLSDID----- 1976
QY 308 ILQEAQMVIAEKDLKNIKPADGSDVPNPGTTVGGSKQ-----QCSSISGISIR 355
Db 1977 --QENNN--KENEPKETEPDPSQGEPSKQASGYAPKSFHVEDTPVCFSRNSSL3SL5 2031
QY 356 VSMLLDDAENETASILMSGFROMIHMFNTENPDSQAAQQLAAQAAKAGDDSAAL 415
Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGDNEKHSPRNMGGILGED-LTLDL 2080
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449
Db 2081 KQIQRDPDSEHGLSPDSENFENFKAIQEGANSIVSLHQAAAAACLS 2125

RESULT 13
US-08-452-654-2
Sequence 2, Application US/08452654
Patent No. 5691454
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, McKie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452,654
FILING DATE: 25-MAY-1995
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-654-2

Query Match 6.18; Score 138; DB 1; Length 2843;
Best Local Similarity 20.08; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55

```

Db 1659 DLTIESPNEAAGEVGGAGSEFEKRDITPTIGRSTDEAOGKTSSTVITPELDDNKA 1718
QY 56 ERWSILRSVNAALM-----SLADKLGIASSNSSSTSRSGADVDSTTATPTPP 103
Db 1719 EEDGILAEICNSAMPKSHKPFVRVKIMDVOQAASASSAPKNKQDGGKKKPTSPVKP 1778
QY 104 PPTSDDYKTOAQAYDT-----IFT-----STSLADIAQALVSLQDAV--- 141
Db 1779 IPQTEYTRVRKNADSKNNLNAERVFSDNKKQNLKNNKDFNDKLPNNEDVRGSGF 1838
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVGAQITELAKYASDNOAILDSL 191
Db 1839 AFDSPHHYPTIEGTPYCFSRNDSLSLDFDDDDVLSREKAEL-----RKAKENK---ESE 1891
QY 192 GKLTSDLLQTLQSVANNKKAELLKEM---QDNPVVPGKTPAQAQSLVDOTDATAT 247
Db 1892 AKVTS-----HTELTSSNOOSANKTQAIKQPINRGQPKPILO--KOSTFPQSSKDIPIRGAA 1946
QY 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAAKAAIATAKTAQIAEAQKKFPDPS 307
Db 1947 TDEKLQ-----FAIENTPVCFSHNSLSLSDID----- 1976
QY 308 ILQEAQOMVIOAEKDLKNIPADGSDVPNPGTTVGGSKQ-----QGSSIGSIR 355
Db 1977 --QENNN---KENEPIKETEPDPSQGEPSKPAQSGYAPKSFHVEDTPVCFSRNSSLSSLS 2031
QY 356 VSMLLDDAENETASILMSGFQMIHMENTENPDQQAQELAAQAAKAAAGDDSAAL 415
Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGDNEKHSRNNGGILGED-LTLDL 2080
QY 416 ADAQKA-LEAALG-----KAGQ--QGILNALQIAAAVVS 449
Db 2081 KDIQRPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAACLS 2125

RESULT 14

US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Witcoff, Ltd.
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,655B
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/289,548
; FILING DATE: 12-AUG-1994

PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.49964
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-655B-2

Query Match 6.1%; Score 138; DB 1; Length 2843;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;
QY 11 DETERPPADLSA-QGLEASAANKSARAQRIAGAEAKPK-----SKTDSV----- 55
Db 1659 DLTIESPNEAAGEVGGAGSEFEKRDITPTIGRSTDEAOGKTSSTVITPELDDNKA 1718
QY 56 ERWSILRSVNAALM-----SLADKLGIASSNSSSTSRSGADVDSTTATPTPP 103
Db 1719 EEDGILAEICNSAMPKSHKPFVRVKIMDVOQAASASSAPKNKQDGGKKKPTSPVKP 1778
QY 104 PPTSDDYKTOAQAYDT-----IFT-----STSLADIAQALVSLQDAV--- 141
Db 1779 IPQTEYTRVRKNADSKNNLNAERVFSDNKKQNLKNNKDFNDKLPNNEDVRGSGF 1838
QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVGAQITELAKYASDNOAILDSL 191
Db 1839 AFDSPHHYPTIEGTPYCFSRNDSLSLDFDDDDVLSREKAEL-----RKAKENK---ESE 1891
QY 192 GKLTSDLLQTLQSVANNKKAELLKEM---QDNPVVPGKTPAQAQSLVDOTDATAT 247
Db 1892 AKVTS-----HTELTSSNOOSANKTQAIKQPINRGQPKPILO--KOSTFPQSSKDIPIRGAA 1946
QY 248 QIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAAKAAIATAKTAQIAEAQKKFPDPS 307
Db 1947 TDEKLQ-----FAIENTPVCFSHNSLSLSDID----- 1976
QY 308 ILQEAQOMVIOAEKDLKNIPADGSDVPNPGTTVGGSKQ-----QGSSIGSIR 355
Db 1977 --QENNN---KENEPIKETEPDPSQGEPSKPAQSGYAPKSFHVEDTPVCFSRNSSLSSLS 2031
QY 356 VSMLLDDAENETASILMSGFQMIHMENTENPDQQAQELAAQAAKAAAGDDSAAL 415
Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGDNEKHSRNNGGILGED-LTLDL 2080
QY 416 ADAQKA-LEAALG-----KAGQ--QGILNALQIAAAVVS 449
Db 2081 KDIQRPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAACLS 2125

RESULT 15

US-08-452-655B-7
; Sequence 7, Application US/08452655B
; Patent No. 5783666
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
 NUMBER OF SEQUENCES: 102
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Banner & Witcoff, Ltd.
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA
 ZIP: 20001-4598
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,655B
 FILING DATE: 25-MAY-1995
 CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/289,548
 FILING DATE: 12-AUG-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.49964
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299
 INFORMATION FOR SEQ ID NO: 7:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2843 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: YES
 ANTI-SENSE: NO
 US-08-452-655B-7

Query Match 6.1%; Score 138; DB 1; Length 2843;
 Best Local Similarity 20.0%; Pred. No. 0.11;
 Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERPPADLSA-OGLEASANKSAEAOIAGAEAKPE-----SKTDSV----- 55
 Db 1659 DLTIESPPNELAGEGVGGAQSGEKEKRDITETGRSTDEAOGGKTSVVTPELDDNKA 1718
 QY 56 ERWSILRSVNAALM-----SLADKLGIASNSSSTSRSDVDSTTATPTPP 103
 Db 1719 EEGDILAEACINSAMPKSHKPRVKKIMDQVOOASASSAPNKNQLDGKKKPTSPVKP 1778
 QY 104 PPTSDYKTKQAQAYDT-----IFT-----STSLADIQAALVSLQDAV---- 141
 Db 1779 IPONTEYRTRVRKNADSKNLNAERVFSDNKSQKKNLNNKSKDFNDKLPNNEDVRVRSF 1838
 QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191
 Db 1839 AFDSPHHYPIETGTCFSDNSLSDPDVDDVLSREKAE-----RKAKENK-----ESE 1891
 QY 192 GKLTSEDLQATALLQSVANNKAAELKEM-----QDNVVPVGTPTPAQSLVDQTATAT 247
 Db 1892 AKVTS-----HTELTNQOSANKTQAIKQPINRGQPKPIQ-KQSTFPQSSKDIPDRGAA 1946
 QY 248 QIEKDGNAIGDAYFAGONAGVAKSNNSNISNIDSAKAAIATAKTQIAEAQKFPDPS 307
 Db 1947 TDEKLQN-----FAIENTVPCFSHNSLSLSDID----- 1976

QY 308 ILQEAQWVIAQEKDLKNIKPADGSDVPNPGTIVGSKQ-----QGSSTGSGIR 355
 Db 1977 --QENN---KENEPKETEPDPSQGEPSKPOASGVAPKSEHVEDTPVCFSRNSSLSLS 2031
 QY 356 VSMLLDDAENETASIIIMSGFROMIHMFNTENPDSSQAQOELAAQARAAGDSDSANAAL 415
 Db 2032 I-----DSEDDLLQECISS-----AMPKKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2080
 QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALQOIASAAVVS 449
 Db 2081 KDIQRPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAACLS 2125

Search completed: February 7, 2002, 21:36:24
 Job time: 20493 sec

Result No.	Query			Length	DB	ID	Description
	Score	Match					
1	1947	92.6	389	2	A43587	major	outer membra
2	1947	92.6	389	2	D86577	major	outer membra
3	1310	90.8	389	2	I40864	major	outer membra
4	1882	89.5	389	2	I40739	major	outer membra
5	1534.5	73.0	389	1	MMCWP3	major	outer membra
6	1496.5	71.2	389	2	A60109	major	outer membra
7	1477	70.2	392	2	A40371	major	outer membra
8	1419	67.5	402	2	I40740	major	outer membra
9	1415	67.3	402	1	MMCWPM	major	outer membra
10	1410	67.0	402	2	B60109	major	outer membra
11	1408	67.0	402	2	A60341	major	outer membra
12	1351.5	64.3	387	2	C81747	major	outer membra
13	1351	64.2	404	2	I40741	major	outer membra
14	1349.5	64.2	387	2	JT0947	mouse	pneumonitis
15	1349.5	64.2	387	2	S16034	major	outer membra
16	1340.5	63.7	397	1	MMCWTH	major	outer membra
17	1338.5	63.6	397	2	JF0413	major	outer membra
18	1333.5	63.4	375	2	S11007	major	outer membra
19	1328	63.1	374	2	S11006	major	outer membra
20	1325	63.0	396	2	S12799	major	outer membra
21	1315.5	62.6	393	1	MMCWTE	major	outer membra
22	1315.5	62.6	397	1	MMCWTC	major	outer membra
23	1313.5	62.5	393	2	S06259	major	outer membra
24	1310.5	62.3	393	2	T01645	major	outer membra
25	1308.5	62.2	393	2	H71484	probable	major out
26	1307.5	62.2	395	1	MMCWTF	major	outer membra
27	1303.5	62.0	393	2	JC1432	major	outer membra
28	1295	61.6	394	2	S11012	major	outer membra
29	1292	61.4	372	2	B60759	major	outer membra

A43587
major outer membrane protein, porin CP0051 precursor [imported] - Chlamydomophila pneumoniae
N:Alternate names: MOMP
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 29-Jan-1993 #sequence.revision 29-Jan-1993 #text_change 11-May-2000
C:Accession: A43587; A49751; A49216; G72044; F81619
C:Peretz Melgosa, M.; Kuo, C.C.; Campbell, L.A.
Infect. Immun. 59, 2195-2199, 1991
A:Title: Sequence analysis of the major outer membrane protein gene of Chlamydia pneumoniae
A:Reference number: A43587; MUID:91244474
A:Accession: A43587
A:Molecule type: DNA
A:Residues: 1-389 <PER>
A:Cross-references: GB:M69230; NID:g144540; PIDN:AAA73071.1; PID:g144541
R:Carter, M.W.; Al-Mahdawi, S.A.H.; Gilles, I.G.; Treharne, J.D.; Ward, M.E.; Clarke, T.
J. Gen. Microbiol. 137, 465-475, 1991
A:Title: Nucleotide sequence and taxonomic value of the major outer membrane protein gene of Chlamydia pneumoniae
A:Reference number: A49751; MUID:91237311
A:Accession: A49751
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <CAR>
A:Cross-references: GB:M64064; GB:M34942; NID:g144534; PIDN:AAA23143.1; PID:g144535
R:Gaydos, C.A.; Quinn, T.C.; Bobo, L.D.; Eiden, J.J.
Infect. Immun. 60, 5319-5323, 1992
A:Title: Similarity of Chlamydia pneumoniae strains in the variable domain IV region
A:Reference number: A49216; MUID:93084388
A:Accession: A49216
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 297-352 <GAY>
A:Cross-references: GB:S06072; NID:g260972; PIDN:AAB24363.1; PID:g260973
A:Note: sequence extracted from NCBI backbone (NCBIN:120604, NCBI:P:120605)
R:Kallan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, N.
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: G72044
A:Molecule type: DNA
A:Residues: 1-389 <ARN>
A:Cross-references: GB:AE001652; GB:AE001363; NID:g4376997; PIDN:AAD18834.1; PID:g4376998
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickman, C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L.; et al.
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR3
A:Reference number: AB1500; MUID:20150255
A:Accession: F81619
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-389 <REA>
A:Cross-references: GB:AE002168; GB:AE002161; NID:g7188982; PIDN:AAF37944.1; PID:g718899
A:Experimental source: strain AR39, HL cells
C:Genetics:

C:Gene: ompA; CP0051
C:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 92.6%; Score 1947; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.7e-148;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 2 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 24 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
QY 62 SMGAKPTGSAANTTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 121
DB 84 SMGAKPTGSAANTTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 181
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 203
QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSATI 241
DB 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSATI 263
QY 242 NYHEQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
DB 264 NYHEQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361
DB 324 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 383
QY 362 SQQFRF 367
DB 384 SQQFRF 389
```

RESULT 2

D86577
major outer membrane protein [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86577
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: D86577
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STC>
A:Cross-references: GB:BA000008; NID:g8979067; PIDN:BA98902.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: ompA
C:Superfamily: Chlamydia major outer membrane protein

Query Match 92.6%; Score 1947; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.7e-148;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 2 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 24 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
```

```
QY 62 SMGAKPTGSAANTTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 121
DB 84 SMGAKPTGSAANTTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 181
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 203
QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSATI 241
DB 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSATI 263
QY 242 NYHEQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
DB 264 NYHEQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361
DB 324 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 383
QY 362 SQQFRF 367
DB 384 SQQFRF 389
```

RESULT 3

I40864

major outer membrane protein - Chlamydia psittaci
C:Species: Chlamydia psittaci, Chlamydia psittaci
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
C:Accession: I40864; S33465
R:Gijses, A.A.; Carrick, F.N.; Lavin, M.F.
Gene 138, 139-142, 1994
A:Title: Remarkable sequence relatedness in the DNA encoding the major outer membra
A:Reference number: I40864; MUID:94171025
A:Accession: I40864
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-389 <RES>
A:Cross-references: EMBL:X72023; NID:g313844; PIDN:CAA50906.1; PID:g313845
C:Superfamily: Chlamydia major outer membrane protein

Query Match 90.8%; Score 1910; DB 2; Length 389;
Best Local Similarity 97.5%; Pred. No. 2.5e-145;
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
QY 2 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 24 LPVGNPSPDLLIDGTIWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
QY 62 SMGAKPTGSAANTTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 121
DB 84 SMGAKPTGSAANTTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDFVCTLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 181
DB 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTDFVCTLGASNG 203
QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSATI 241
DB 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGKTSATI 263
QY 242 NYHEQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
DB 264 NYHEQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 361
DB 324 TTTLATSDSFDMQIVSCQINKFKSRKACGVTGATLVADKWSLTAEARLINERAHV 383
QY 362 SQQFRF 367
DB 384 SQQFRF 389
```


Db 384 SQQFRF 389

RESULT 4

140739 major outer membrane protein precursor - Chlamydomydia pneumoniae (strain equine/N16)
 C:Species: Chlamydomydia pneumoniae, Chlamydia pneumoniae
 A:Variety: strain equine/N16
 C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 20-Apr-2000
 C:Accession: I40739
 R:Storey, C.; Lusher, M.; Yates, P.; Richmond, S.
 J. Gen. Microbiol. 139, 2621-2626, 1993
 A:Title: Evidence for Chlamydia pneumoniae of non-human origin.
 A:Reference number: I40739; MUID:94103736
 A:Accession: I40739
 A:Status: translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-389 <STO>
 A:Cross-references: GB:L04982; NID:g289840; PIDN:AA17397.1; PID:g289841
 C:Comment: On the basis of the major outer membrane protein the authors classified the
 the sequence of the genome strain CWL029 and strain IOL-207. See PIR:A43587.
 C:Genetics:
 A:Gene: momp
 C:Superfamily: Chlamydia major outer membrane protein
 C:Keywords: membrane protein
 F:1-23/Domain: signal sequence #status predicted <SIG>
 F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 89.5%; Score 1882; DB 2; Length 389;

Best Local Similarity 94.5%; Pred. No. 4.3e-143;

Matches 346; Conservative 15; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPVGNPSPSLIDGTIWEGAAGDPCPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 24 LPVGNPSPSLIDGTIWEGAAGDPCPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 83

QY 62 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEWFTNAGFIATLNIWDRFVFCITLGASNG 121

DB 84 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEWFTNAGFIATLNIWDRFVFCITLGASNG 143

QY 122 YIRGNSTAFNLVGLFGVKGTTVNAELPNVSLNSGVVLYTDTSPFSWSVGARGALWECG 181

DB 144 YVKGNSAFLNLVGLFGVKGTTVNAELPNVSLNSGVVLYTDTSPFSWSVGARGALWECG 203

QY 182 ATLGAEFYQAQSKPVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 241

DB 204 ATLGAEFYQAQSKPVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKTSATI 263

QY 242 NYHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301

DB 264 NYHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGS 323

QY 302 ATALSTTDSFDMQIVSCQINKFSKACGVTGATLVDAKWSLTAEARLINERAHV 361

DB 324 ATAVSSDDQFDMQIVSCQINKFSKACGVTGATLVDAKWSLTAEARLINERAHI 383

QY 362 SQQFRF 367

DB 384 SQQFRF 389

RESULT 5

MMCW3 major outer membrane protein precursor - Chlamydomydia psittaci (strain S26/3)
 C:Species: Chlamydomydia psittaci, Chlamydia psittaci
 C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 31-Mar-2000
 C:Accession: S08770
 R:Herring, A.J.; Tan, T.W.; Baxter, S.; Inglis, N.F.; Dunbar, S.
 FEMS Microbiol. Lett. 65, 153-158, 1989
 A:Title: Sequence analysis of the major outer membrane protein gene of an ovine abortion
 A:Reference number: S08770
 A:Accession: S08770

A:Molecule type: DNA

A:Residues: 1-389 <HER>

A:Cross-references: EMBL:X51859; NID:g40600; PIDN:CAA36152.1; PID:g40601

C:Superfamily: Chlamydia major outer membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 73.08; Score 1534.5; DB 1; Length 389;

Best Local Similarity 75.7%; Pred. No. 2.9e-115;

Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPSPSLIDGTIWEGAAGDPCPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 23 LPVGNPAEPLIDGTIWEGASGDCPCSTWCDALSLRAGFYGYVDFRILKVDVNTKI 82

QY 62 S-MGAKPTGSAANYTTAVDRPNPAYNKLHDAEWFTNAGFIATLNIWDRFVFCITLGASN 120

DB 83 TCMGAVPTGTAANYKTPTDRPNIAYGKHLQDAEWFTNAAFLALNIWDRFDFICTLGASN 142

QY 121 GYIRGNSTAFNLVGLFGVKGTTVNAELPNVSLNSGVVLYTDTSPFSWSVGARGALWECG 180

DB 143 GYFKAASAFNLVGLFGVKGTTVNAELPNVSLNSGVVLYTDTSPFSWSVGARGALWECG 202

QY 181 CATLGAEFYQAQSKPVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKTSAT 240

DB 203 CATLGAEFYQAQSKPVEELNVCNVSQFSVKNPKYKGVAFPLPTDAGVATATGKTSAT 262

QY 241 INYHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLG 300

DB 263 IKYHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLG 322

QY 301 NATALSTTDSFDMQIVSCQINKFSKACGVTGATLVDAKWSLTAEARLINERAHV 360

DB 323 EATALDTSNKAFLQIASIQINKMKSRKACGVAATLVDAKWSLTAEARLINERAHV 382

QY 361 VSGQFRF 367

DB 383 MNAQFRF 389

RESULT 6

A60109 major outer membrane protein precursor - Chlamydomydia psittaci (strain Guinea pig 1
 C:Species: Chlamydomydia psittaci, Chlamydia psittaci
 C:Date: 10-Nov-1992 #sequence_revision 10-Nov-1992 #text_change 31-Mar-2000
 C:Accession: A60109
 R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.
 Infect. Immun. 57, 1621-1625, 1989
 A:Title: Cloning and sequence analysis of the major outer membrane protein genes of
 A:Reference number: A60109; MUID:89212917
 A:Accession: A60109
 A:Status: not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-389 <ZHA>
 C:Superfamily: Chlamydia major outer membrane protein
 C:Keywords: membrane protein
 F:1-22/Domain: signal sequence #status predicted <SIG>
 F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 71.2%; Score 1496.5; DB 2; Length 389;

Best Local Similarity 73.9%; Pred. No. 3.2e-112;

Matches 272; Conservative 42; Mismatches 51; Indels 3; Gaps 2;

QY 2 LPVGNPSPSLIDGTIWEGAAGDPCPCATWCDALSLRAGFYGYVDFRILKVDPAKTF 61

DB 23 LPVGNPAEPLIDGTIWEGASGDCPCSTWCDALSLRAGFYGYVDFRILKVDVNTKI 82

QY 62 SMGAKPTGSAANYTTAVDRPNPAYNKLHDAEWFTNAGFIATLNIWDRFVFCITLGASN 121

DB 83 SMGTAPTGNAADFTVADRNNIAYGKMQDAEWSTNAAFLALNIWDRFVFCITLGASN 142

F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-402/Product: major outer membrane protein #status predicted <MAT>

Query Match 67.3%; Score 1415; DB 1; Length 402;
Best Local Similarity 70.3%; Pred. No. 1.le-105;
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPSPDLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDPAKTF 61
DB 23 LPVGNPAEPSSLIDGTWEGASGDCPCDPCATWCDALSLRAGFYGDYVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLEDAEFTNAGFIALNIWDRFVFTCL 116
DB 83 SCMAATPTQATGNASNTQPEANGRPNTAYGRHMQDAEWFNAFLALNIWDRFDFICTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKGTIVANE-----LPNVLSNGVVELYTDTSFWSVGA 172
DB 143 GASNGYFKSSAFLNLVGLIGFSSTSTSLPMLPNVITQGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEOAQSFKPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 230
DB 203 RGALWECGCATLGAEOAQSFKPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVGLSRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
DB 263 TEATDTKTSATIKYHEWQVGLSRLNSLPVYIGVQWSRATFDADTIRIAQPKLSEILN 322

QY 291 LTAWNPSLLGNATALSTT---DSFSDFMQIVSCQINKKFSKACGVTVGATLVADAKWSL 347
DB 323 ITTWNPSLLGSTTTLPNNGCKDVLSDVLOIASIQINKMKSRKACGVAVGATLIDAKWSI 382

QY 348 TAEARLINERAHVHVSQGRF 367
DB 383 TGEARLINERAHMAHQRF 402

RESULT 10
B60109
major outer membrane protein precursor - Chlamydomophila psittaci (strain meningopneumoni
C:Species: Chlamydomophila psittaci, Chlamydia psittaci
C:Date: 10-Nov-1992 #sequence_revision 10-Nov-1992 #text_change 31-Mar-2000
C:Accession: B60109
R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.
Infect. Immun. 57, 1621-1625, 1989
A:Title: Cloning and sequence analysis of the major outer membrane protein genes of two
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-389/Product: major outer membrane protein #status predicted <MAT>

A:Reference number: A60109; MUID:89212917
A:Accession: B60109
A:Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-402 <ZHA>
C:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 67.0%; Score 1410; DB 2; Length 402;
Best Local Similarity 70.3%; Pred. No. 2.8e-105;
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSPDLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDPAKTF 61
DB 23 LPVGNPAEPSSLIDGTWEGASGDCPCDPCATWCDALSLRAGFYGDYVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLEDAEFTNAGFIALNIWDRFVFTCL 116
DB 83 SCMAATPTQATGNASNTQPEANGRPNTAYGRHMEDAEWFNAFLALNIWDRFDFICTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG-----TTVNANELPNVLSNGVVELYTDTSFWSVGA 172
DB 143 GASNGYFKSSAFLNLVGLIGFSSTSTSLPMLPNVITQGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEOAQSFKPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 230
DB 203 RGALWECGCATLGAEOAQSFKPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVGLSRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
DB 263 TEATDTKTSATIKYHEWQVGLSRLNSLPVYIGVQWSRATFDADTIRIAQPKLSEILN 322

QY 291 LTAWNPSLLGNATAL---STTDSFSDFMQIVSCQINKKFSKACGVTVGATLVADAKWSL 347
DB 323 ITTWNPSLLGSTTTLPNNGCKDVLSDVLOIASIQINKMKSRKACGVAVGATLIDAKWSI 382

QY 348 TAEARLINERAHVHVSQGRF 367
DB 383 TGEARLINERAHMAHQRF 402

RESULT 11

A60341
major outer membrane protein precursor - Chlamydomophila psittaci (strain 6BC)
C:Species: Chlamydomophila psittaci, Chlamydia psittaci
C:Date: 11-Dec-1992 #sequence_revision 24-Feb-1994 #text_change 31-Mar-2000
C:Accession: A44565; A60341; B60341
R:Everett, K.D.E.
submitted to the EMBL Data Library, December 1990
A:Reference number: A44565
A:Accession: A44565
A:Molecule type: DNA
A:Residues: 1-402 <EVE>
A:Cross-references: GB:X56980; NID:g40568; PIDN:CAA40300.1; PID:g40569
R:Everett, K.D.E.; Andersen, A.A.; Plaunt, M.; Hatch, T.P.
Infect. Immun. 59, 2853-2855, 1991
A:Title: Cloning and sequence analysis of the major outer membrane protein gene of
A:Reference number: A60341; MUID:91310346
A:Accession: A60341
A:Molecule type: protein
A:Residues: 23-35 <EV2>
A:Accession: B60341
A:Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 112-232; 317-349 <EV3>
A:Cross-references: GB:X56980
C:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-402/Product: major outer membrane protein #status experimental <MAT>

Query Match 67.0%; Score 1408; DB 2; Length 402;
Best Local Similarity 70.3%; Pred. No. 4e-105;
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSPDLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDPAKTF 61
DB 23 LPVGNPAEPSSLIDGTWEGASGDCPCDPCATWCDALSLRAGFYGDYVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLEDAEFTNAGFIALNIWDRFVFTCL 116
DB 83 SCMAATPTQATGNASNTQPEANGRPNTAYGRHMQDAEWFNAFLALNIWDRFDFICTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG-----TTVNANELPNVLSNGVVELYTDTSFWSVGA 172
DB 143 GASNGYFKSSAFLNLVGLIGFSSTSTSLPMLPNVITQGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEOAQSFKPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 230
DB 203 RGALWECGCATLGAEOAQSFKPKVEELNVCNVSQFVSNKPKGYKGA--FPLPTDAGV 262

QY 231 ATATGKTSATINHEWQVGLSRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
DB 263 TEATDTKTSATIKYHEWQVGLSRLNSLPVYIGVQWSRATFDADTIRIAQPKLSEILN 322

QY 291 LTAWNPSLLGNATAL---STTDSFSDFMQIVSCQINKKFSKACGVTVGATLVADAKWSL 347

QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFTLGA 118
 DB 83 EMGAAPTGD--ADLTAPTASRENPAKQKHQDAEMFTNAYMALNIWDRFDVFTLGA 140
 QY 119 SNGYIRGNSTANLVGLFGVKTNNANLPLNVSLSNGVVELYTDTSFWSVGARGALWE 178
 DB 141 TSGYLKGNAAAFNLVGLFRDETAADIPNVLSQAQVVELYTDTSFWSVGARGALWE 200
 QY 179 CGCATLGAEFOYQAQSKPKVEELNVCNVSQFVSNKPKGYKGVAFPLPTDAGVATATGKS 238
 DB 201 CGCPTLGASFQAQSKPKVEELNVCNVAEFTNPKGYGQEFPLNKGAGTVSATDTKD 260
 QY 239 ATINHEWQVGSLSYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298
 DB 261 ASIDYHEWQASLSYRLNMFYPIGVKWSRASFDADTIRIAQPKLETSILKMTWNPPTI 320
 QY 299 LGNATALSTDSFSDFMQIVSCQINKFKSRKACGVTVGATLVADADKWSLTAEARLINERA 358
 DB 321 SSGSIDVDVT--KITDTLQIVSLQNLNKMKSRCGLAIGTTIVDADKYAVTVETRLIDERA 378
 QY 359 AHVSGQFRF 367
 DB 379 AHVNAQFRF 387

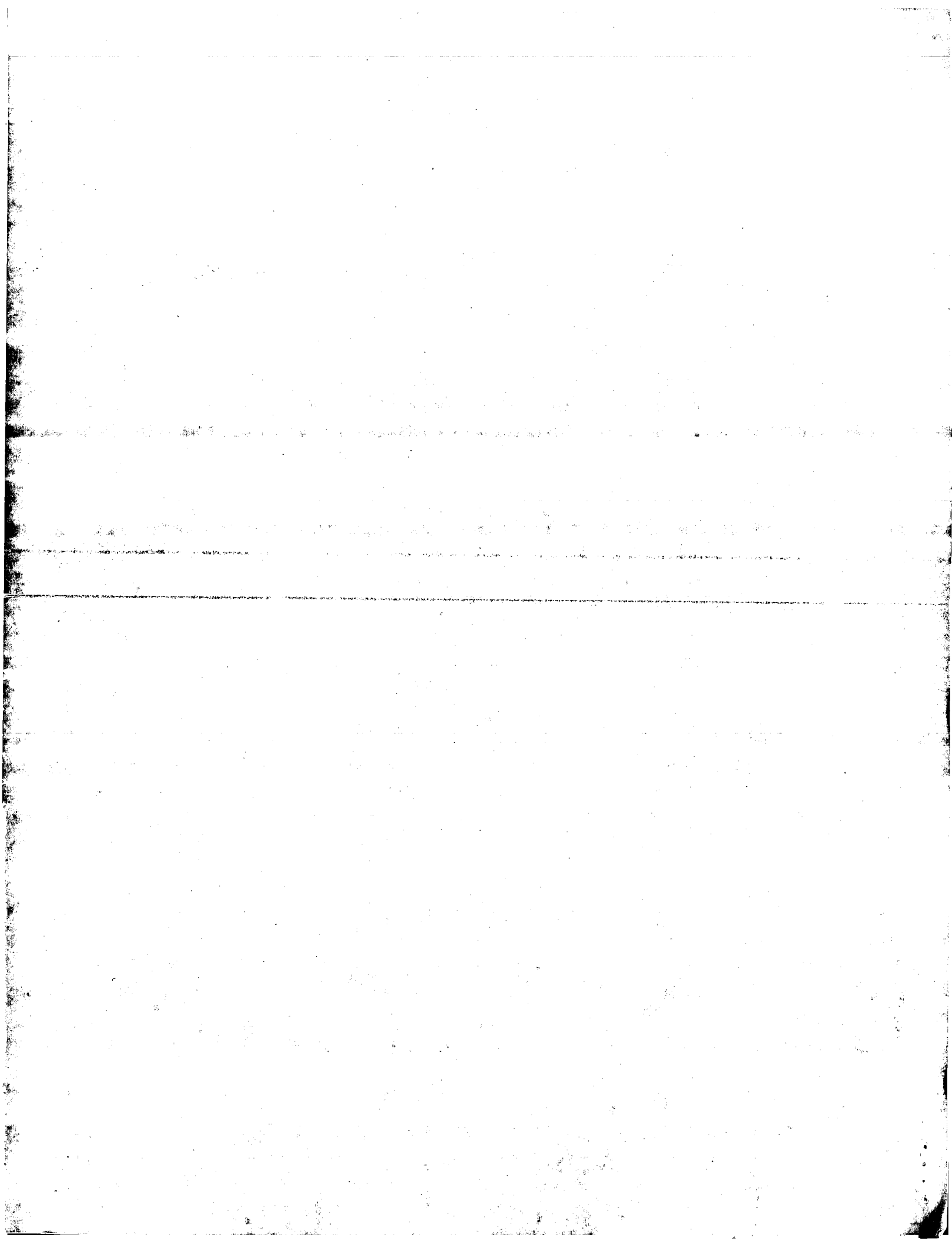
RESULT 15

SI6034
 major outer membrane protein -- Chlamydia trachomatis
 C:Species: Chlamydia trachomatis
 C:Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 12-Apr-1996
 C:Accession: SI6034; A43875
 R:Zhang, Y.X.; Zhang, L.
 submitted to the EMBL Data Library, July 1991
 A:Description: The nucleotide sequence of major outer membrane protein gene of mouse bio
 A:Reference number: SI6034
 A:Accession: SI6034
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-387 <ZHA>
 A:Cross-references: EMBL:X60678
 R:Peterson, E.M.; Cheng, X.; Markoff, B.A.; Fielder, T.J.; de la Maza, L.M.
 Infect. Immun. 59, 4147-4153, 1991
 A:Title: Functional and structural mapping of Chlamydia trachomatis species-specific ma
 A:Reference number: A43875; MUID:92040090
 A:Accession: A43875
 A>Status: Preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 314-322 <PET>
 A:Experimental source: mouse pneumonitis strain
 A:Note: sequence extracted from NCBI backbone (NCBIP:62877)
 C:Superfamily: Chlamydia major outer membrane protein

Query Match 64.2%; Score 1349.5; DB 2; Length 387;
 Best Local Similarity 66.1%; Pred. No. 1.9e-100;
 Matches 244; Conservative 50; Mismatches 68; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTINWGAAGDPCDCAITWCDALISLRAFYGVDYFDRILKVDAPKTF 61
 DB 23 LPVGNPAEPLSLIDGILWEGFGDPCDCTTWCDAISLRLGYGDFVDRVLKTDVNKQF 82
 QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFTLGA 118
 DB 83 EMGAAPTGD--ADLTAPTASRENPAKQKHQDAEMFTNAYMALNIWDRFDVFTLGA 140
 QY 119 SNGYIRGNSTANLVGLFGVKTNNANLPLNVSLSNGVVELYTDTSFWSVGARGALWE 178
 DB 141 TSGYLKGNAAAFNLVGLFRDETAADIPNVLSQAQVVELYTDTSFWSVGARGALWE 200
 QY 179 CGCATLGAEFOYQAQSKPKVEELNVCNVSQFVSNKPKGYKGVAFPLPTDAGVATATGKS 238
 DB 201 CGCATLGASFQAQSKPKVEELNVCNVAEFTNPKGYGQEFPLNKGAGTVSATDTKD 260

Search completed: February 7, 2002, 21:38:10
 Job time: 7099 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:45 ; Search time 76.51 seconds
(without alignments)
188.811 Million cell updates/sec

Title: US-09-391-606-15

Perfect score: 2103

Sequence: 1 MLPVGNPSPDLLIDTWE.....QKLISEDLNSAVDHHHHH 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	92.6	389	1	OMPL_CHLPN P27455 chlamydia p
2	1882	89.5	389	1	OMIN_CHLPN Q07430 chlamydia p
3	1732	82.4	333	1	OMIK_CHLPN Q9XBf4 chlamydia p
4	1534.5	73.0	389	1	OMIA_CHLPS P16567 chlamydia p
5	1477	70.2	392	1	OMIP_CHLPS Q00087 chlamydia p
6	1415	67.3	402	1	OMLE_CHLPS P10332 chlamydia p
7	1358.5	64.6	387	1	OMPL_CHLMU P75024 chlamydia m
8	1340.5	63.7	397	1	OMIH_CHLTR P13467 chlamydia t
9	1338.5	63.6	397	1	OMIN_CHLTR P23114 chlamydia t
10	1325	63.0	396	1	OMIA_CHLTR P23732 chlamydia t
11	1315.5	62.6	393	1	OMIE_CHLTR P17451 chlamydia t
12	1315.5	62.6	397	1	OMIC_CHLTR P08780 chlamydia t
13	1313.5	62.5	393	1	OMIL_CHLTR P19542 chlamydia t
14	1308.5	62.2	393	1	OMID_CHLTR Q46409 chlamydia t
15	1307.5	62.2	395	1	OMIF_CHLTR P16155 chlamydia t
16	1295	61.6	394	1	OMIM_CHLTR P06597 chlamydia t
17	1275	60.6	394	1	OMIB_CHLTR P23421 chlamydia t
18	189	9.0	340	1	OMP2_CHLTR P38006 chlamydia t
19	182	8.7	340	1	OMP2_CHLMU Q9p113 chlamydia m
20	147.5	7.0	344	1	OMP2_CHLPN Q92752 chlamydia p
21	103	4.9	864	1	AGLU_MUCJA Q92442 mucor javan
22	100.5	4.8	919	1	HEX_ADEL2 P19900 human adeno
23	98	4.7	1848	1	CPBA_CLOCL P38058 clostridium
24	96	4.6	556	1	PHL_LEPIN P17627 leptospira
25	96	4.6	3421	1	TEGU_HSVBP P28955 equine herp
26	94	4.5	448	1	FADL_ECOLI P10384 escherichia
27	93.5	4.4	2021	1	OMPA_RICCN Q52657 rickettsia
28	93.5	4.4	2292	1	POLG_EMCVB P17593 encephalomy
29	93.5	4.4	2292	1	POLG_EMCVD P17594 encephalomy
30	92.5	4.4	1169	1	YK82_YEAST P36170 saccharomyc
31	92.5	4.4	2290	1	POLG_EMCV P03304 encephalomy
32	91.5	4.4	639	1	AMVG_ASPSH P22832 aspergillus
33	90.5	4.3	1086	1	NTM_MOUSE Q61941 mus musculu

RESULT 1

ID	OMPL_CHLPN	STANDARD;	PRT;	389 AA.
AC	P27455; Q9JQF6;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).			
GN	OMPA OR OMPI OR CPN0695 OR CP0051.			
OS	Chlamydia pneumoniae (Chlamydophila pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IOL-207;			
RX	MEDLINE=91237311; PubMed=2033374;			
RA	Carter M.W., Al-Mahdawi S.A.H., Giles I.G., Trehan J.D.,			
RA	Ward M.E., Clarke I.N.;			
RT	"Nucleotide sequence and taxonomic value of the major outer membrane			
RT	protein gene of Chlamydia pneumoniae IOL-207.";			
RL	J. Gen. Microbiol. 137:465-475(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TWAR;			
RX	MEDLINE=91244474; PubMed=1840574;			
RA	Perez Melgosa M., Kuo C.-C., Campbell L.A.;			
RT	"Sequence analysis of the major outer membrane protein gene of			
RT	Chlamydia pneumoniae.";			
RL	Infect. Immun. 59:2195-2199(1991).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Mitchell W.M., Tharp A.C., Stratton C.W., Sriram S.;			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=CHL029;			
RX	MEDLINE=92026606; PubMed=10192388;			
RA	Kalman S., Grinchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Olinger L., Mitchell J., Davis R.W., Stephens R.S.;			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";			
RL	Nat. Genet. 21:385-389(1999).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AR39;			
RX	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,			
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RA	Eisen J., Fraser C.M.;			
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			
RT	pneumoniae AR39.";			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=J138;			

34	90.5	4.3	1140	1	YHJL_ECOLI	P37650 escherichia
35	90	4.3	533	1	FCY2_YEAST	P17064 saccharomyc
36	89.5	4.3	726	1	FATA_VIBAN	P11461 vibrio angu
37	89.5	4.3	867	1	SEMD_ECOLI	P77468 escherichia
38	89.5	4.3	901	1	POLG_ENMG3	P32540 mengo encep
39	89	4.2	208	1	Y396_RICPR	Q92dd5 rickettsia
40	89	4.2	363	1	YRAK_ECOLI	P43319 escherichia
41	89	4.2	895	1	Z281_HUMAN	O9V2x9 homo sapien
42	88.5	4.2	561	1	HLVB_PROMI	P16465 proteus mir
43	88.5	4.2	639	1	AMYG_ASPAK	P23176 aspergillus
44	88.5	4.2	1035	1	MT10_YEAST	P39692 saccharomyc
45	87.5	4.2	482	1	PUR8_YEAST	Q05911 saccharomyc

ALIGNMENTS

Query Match	92.6%	Score 1947;	DB 1;	Length 389;
Best Local Similarity	100.0%;	pred. No. 1.6e-153;		
Matches 366:	Conservative	0;	Mismatches	0;
	Indels	0;		

QY	242	NYHEQVQASLSYRLNSLVPYIGVQMSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGN	301																																																																																																																								
Db	264	NYHEQVQASLSYRLNSLVPYIGVQMSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLGN	323																																																																																																																								
QY	302	ATALSTTDSFDFMQIVSCQINFKSRKACGVTVGATLVADAKWSLTAEARLINEAAHV	361																																																																																																																								
Db	324	ATALSTTDSFDFMQIVSCQINFKSRKACGVTVGATLVADAKWSLTAEARLINEAAHV	383																																																																																																																								
QY	362	SGQFRF 367																																																																																																																									
Db	384	SGQFRF 389																																																																																																																									
<p>RESULT 2</p> <table border="0"> <tr> <td>OMIN_CHLPN</td><td>STANDARD;</td><td>PRT;</td><td>389 AA.</td></tr> <tr> <td>IC</td><td>Q07430;</td><td></td><td></td></tr> <tr> <td>DT</td><td>30-MAY-2000 (Rel. 39, Created)</td><td></td><td></td></tr> <tr> <td>DT</td><td>30-MAY-2000 (Rel. 39, Last sequence update)</td><td></td><td></td></tr> <tr> <td>DT</td><td>20-AUG-2001: (Rel. 40, Last annotation update)</td><td></td><td></td></tr> <tr> <td>DE</td><td>MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).</td><td></td><td></td></tr> <tr> <td>GN</td><td>OMPA OR OMP1.</td><td></td><td></td></tr> <tr> <td>OS</td><td>Chlamydia pneumoniae (Chlamydia pneumoniae).</td><td></td><td></td></tr> <tr> <td>OC</td><td>Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.</td><td></td><td></td></tr> <tr> <td>NCBI_TaxID=83558;</td><td></td><td></td><td></td></tr> <tr> <td>ON</td><td>[1]</td><td></td><td></td></tr> <tr> <td>RP</td><td>SEQUENCE FROM N.A.</td><td></td><td></td></tr> <tr> <td>RC</td><td>STRAIN=N16;</td><td></td><td></td></tr> <tr> <td>RX</td><td>MEDLINE=94103736; PubMed=8277245;</td><td></td><td></td></tr> <tr> <td>RA</td><td>Storey C., Lusher M., Yates P., Richmond S.;</td><td></td><td></td></tr> <tr> <td>RT</td><td>"Evidence for Chlamydia pneumoniae of non-human origin.";</td><td></td><td></td></tr> <tr> <td>RL</td><td>J. Gen. Microbiol. 139:2621-2626(1993).</td><td></td><td></td></tr> <tr> <td>CC</td><td>-1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.</td><td></td><td></td></tr> <tr> <td>CC</td><td>-1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.</td><td></td><td></td></tr> <tr> <td>CC</td><td>-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.</td><td></td><td></td></tr> <tr> <td>CC</td><td>-1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.</td><td></td><td></td></tr> <tr> <td>CC</td><td colspan="3">-----</td></tr> <tr> <td>CC</td><td colspan="3">This SWISS-PROT entry is copyright. It is produced through a collaboration -</td></tr> <tr> <td>CC</td><td colspan="3">between the Swiss Institute of Bioinformatics and the EMBL outstation -</td></tr> <tr> <td>CC</td><td colspan="3">the European Bioinformatics Institute. There are no restrictions on its</td></tr> <tr> <td>CC</td><td colspan="3">use by non-profit institutions as long as its content is in no way</td></tr> <tr> <td>CC</td><td colspan="3">modified and this statement is not removed. Usage by and for commercial</td></tr> <tr> <td>CC</td><td colspan="3">entities requires a license agreement (See http://www.isb-sib.ch/).</td></tr> <tr> <td>CC</td><td colspan="3">or send an email to license@isb-sib.ch.</td></tr> <tr> <td>CC</td><td colspan="3">-----</td></tr> </table>				OMIN_CHLPN	STANDARD;	PRT;	389 AA.	IC	Q07430;			DT	30-MAY-2000 (Rel. 39, Created)			DT	30-MAY-2000 (Rel. 39, Last sequence update)			DT	20-AUG-2001: (Rel. 40, Last annotation update)			DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).			GN	OMPA OR OMP1.			OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			NCBI_TaxID=83558;				ON	[1]			RP	SEQUENCE FROM N.A.			RC	STRAIN=N16;			RX	MEDLINE=94103736; PubMed=8277245;			RA	Storey C., Lusher M., Yates P., Richmond S.;			RT	"Evidence for Chlamydia pneumoniae of non-human origin.";			RL	J. Gen. Microbiol. 139:2621-2626(1993).			CC	-1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.			CC	-1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.			CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.			CC	-1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.			CC	-----			CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -			CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			CC	the European Bioinformatics Institute. There are no restrictions on its			CC	use by non-profit institutions as long as its content is in no way			CC	modified and this statement is not removed. Usage by and for commercial			CC	entities requires a license agreement (See http://www.isb-sib.ch/).			CC	or send an email to license@isb-sib.ch .			CC	-----		
OMIN_CHLPN	STANDARD;	PRT;	389 AA.																																																																																																																								
IC	Q07430;																																																																																																																										
DT	30-MAY-2000 (Rel. 39, Created)																																																																																																																										
DT	30-MAY-2000 (Rel. 39, Last sequence update)																																																																																																																										
DT	20-AUG-2001: (Rel. 40, Last annotation update)																																																																																																																										
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).																																																																																																																										
GN	OMPA OR OMP1.																																																																																																																										
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).																																																																																																																										
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.																																																																																																																										
NCBI_TaxID=83558;																																																																																																																											
ON	[1]																																																																																																																										
RP	SEQUENCE FROM N.A.																																																																																																																										
RC	STRAIN=N16;																																																																																																																										
RX	MEDLINE=94103736; PubMed=8277245;																																																																																																																										
RA	Storey C., Lusher M., Yates P., Richmond S.;																																																																																																																										
RT	"Evidence for Chlamydia pneumoniae of non-human origin.";																																																																																																																										
RL	J. Gen. Microbiol. 139:2621-2626(1993).																																																																																																																										
CC	-1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.																																																																																																																										
CC	-1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.																																																																																																																										
CC	-1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.																																																																																																																										
CC	-1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.																																																																																																																										
CC	-----																																																																																																																										
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration -																																																																																																																										
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -																																																																																																																										
CC	the European Bioinformatics Institute. There are no restrictions on its																																																																																																																										
CC	use by non-profit institutions as long as its content is in no way																																																																																																																										
CC	modified and this statement is not removed. Usage by and for commercial																																																																																																																										
CC	entities requires a license agreement (See http://www.isb-sib.ch/).																																																																																																																										
CC	or send an email to license@isb-sib.ch .																																																																																																																										
CC	-----																																																																																																																										

Query Match 89.5%; Score 1882; DB 1; Length 389;
Best Local Similarity 94.5%; Pred. No. 3.9e-148;
Matches 346; Conservative 15; Mismatches 5; Indels 0; Gaps 0;

Qy	2	LPVGNPSDPSLLIDGTIWEAGAAGDCPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF	61
Db	24	LPVGNPSDPSLLIDGTIWEAGAAGDCPCDPCATWCDAISLRAGFYGDYVDFRILKIDAPKTF	83
Qy	62	SMGAKPTGSAANYTTAVDRPNPAYNKKHLHDAEWFTNAGFIALNIWDREDVFCFTLGSNG	121
Db	84	SMGAKPTGSAATYTTAVDRPNPAYNKKLYLDAEWFTNAGFIALNIWDREDVFCFTLGSNG	143
Qy	122	YIRGNSTAFNLVGLFGVKGTTVANELPNVLSLNGVVELYTDTSFSWSVGARGALMECCG	181
Db	144	YKVGNSAFLNVLFGVKGTSVANELPNVLSLNGVIELYTDTFAMSVGARGALMECCG	203

QY 182 ATLGAEFOYAQSKPVEELNVICNYSQFSVKNPKGYKGVAEPLPTDAGVATATGKTSATI 241
 DB 204 ATLGAEFOYAQSKPVEELNVICNYSQFSVKNPKGYKGVAEPLPTDAGVATATGKTSATI 263
 QY 242 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
 DB 264 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGS 323
 QY 302 ATALSTTDSFDMQVSCQINKEKSRKACGVTGATLVADKWSLTAEARLINERAHV 361
 DB 324 ATAVSSSQFSDFMIVSCQINKEKSRKACGVTGATLVADKWSLTAEARLINERAHI 383
 QY 362 SQQFRF 367
 DB 384 SQQFRF 389
 RESULT 3
 OMIA_CHLPN STANDARD; PRT; 333 AA.
 AC Q9XB4;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN (MOMP) (FRAGMENT).
 GN OMPA OR OMP1.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiaophila.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KOALA TYPE 1;
 RX MEDLINE=93123168; PubMed=8419295;
 RA Kallenboeck B., Kousoulas K.G., Storz J.;
 RT "Structures of and allelic diversity and relationships among the major
 RT outer membrane protein (ompA) genes of the four chlamydial species.";
 RL J. Bacteriol. 175:487-502(1993).
 CC -|- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
 CC -|- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -|- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: M73038; AAD38210.1;
 DR InterPro; IPR000604; Chlamydia_OMP.
 DR Pfam; PF01308; Chlamydia_OMP; 1.
 DR ProDom; PD001717; Chlamydia_OMP; 1.
 KW Outer membrane; Transmembrane; Porin.
 FT NON_TER 1
 FT NON_TER 333 333
 SQ SEQUENCE 333 AA; 35811 MW; 204604512C4C3B3F CRC64;

Query Match 82.4%; Score 1732; DB 1; Length 333;
 Best Local Similarity 97.3%; Pred. No. 7.8e-136;
 Matches 324; Conservative 4; Mismatches 5; Indels 0; Gaps 0;
 QY 23 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKPTGSAANYTTAVDRP 82
 DB 1 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKPTGSAANYTTAVDRP 60
 QY 83 NPAYNKLHDAEWETNAGFIALNIWDRFDVCTLCASNGYIRGNSTAFNLVGLFCVKGKTS 142

DB 61 NPAYNKLHDAEWETNAGFIALNIWDRFDVCTLCASNGYIRGNSTAFNLVGLFCVKGKTS 120
 QY 143 VNANELPNVLSNGVVELYTDTSFWSVGARGALNECCCATLGAEFQYQAQSKPKVEELNV 202
 DB 121 VAANELPNVLSNGVVELYTDTSFWSVGARGALNECCCATLGAEFQYQAQSKPKVEELNV 180
 QY 203 ICNVQFSQVNRKPKGYKGVAEPLPTDAGVATATGKTSATINYHEQVQASLSYRLNSLPVY 262
 DB 181 ICNVAQFSVNRKPKGYKGVAEPLPTDAGVATATGKTSATINYHEQVQASLSYRLNSLPVY 240
 QY 263 ICVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGNATALSTTDSFDMQVSCQI 322
 DB 241 ICVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGNNTLTPTSDSFDMQVSCQI 300
 QY 323 NKEFKSRKACGVTGATLVADKWSLTAEARLIN 355
 DB 301 NKEFKSRKACGVTGATLVADKWSLTAEARLIN 333

RESULT 4
 OMIA_CHLPS STANDARD; PRT; 389 AA.
 AC P16567;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
 GN OMPA OR OMP1.
 OS Chlamydia psittaci (Chlamydia psittaci).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydiaophila.
 OX NCBI_TaxID=83554;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OVINE ENZOOTIC ABORTION ISOLATE S26/3;
 RX MEDLINE=90128177; PubMed=2612883;
 RA Herring A.J., Tan T.W., Baxter S., Inglis N.F., Dunbar S.;
 RT "Sequence analysis of the major outer membrane protein gene of an
 RT ovine abortion strain of Chlamydia psittaci.";
 RL FEMS Microbiol. Lett. 53:153-158(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BOVINE ABORTION ISOLATE BAL;
 RX MEDLINE=96189695; PubMed=8605581;
 RA Griffiths P.C., Plater J.M., Martin T.C., Hughes S.L.,
 RA Hughes K.J., Hewinson R.G., Dawson M.;
 RT "Epizootic bovine abortion in a dairy herd: characterization of a
 RT Chlamydia psittaci isolate and antibody response.";
 RL Br. Vet. J. 151:683-693(1995).
 CC -|- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
 CC -|- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
 CC -|- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -|- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X51859; CAA36152.1;
 DR EMBL; L39020; AAB02850.1;
 DR PIR; S08770; MWCWP3.
 DR InterPro; IPR000604; Chlamydia_OMP.
 DR Pfam; PF01308; Chlamydia_OMP; 1.
 DR ProDom; PD001717; Chlamydia_OMP; 1.
 KW Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 22

DR	PTR; A40371; A40371.
DR	PIR; S16137; S16137.
DR	InterPro; IPR000604; Chlamydia_OMP.
DR	Pfam; PF01308; Chlamydia_OMP; 1.
DR	ProDom; PD001717; Chlamydia_OMP; 1.
KW	Outer membrane; Transmembrane; Porin; Signal.
FT	SIGNAL 1 22
FT	CHAIN 23 392 MAJOR OUTER MEMBRANE PROTEIN.
SQ	SEQUENCE 392 AA; 42069 MW; 88B3C5D90BBA26DB CRC64;
Query Match 70.2%; Score 1477; DB 1; Length 392;	
Best Local Similarity 74.1%; Pred. No. 1.le-114;	
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps	
QY	2 LPVGNPSDLLIDGTIWEAGAGDPDCPCATWCDAISLRAGFYGDYVFDFIRILKVDAPKTF 61 : : : : : : : : :
DB	23 LPVGNPAEPSLLIDGTIWEAGSGDPDCPCATWCDAISIRAGFYGDYVFDFIRILKYDNKTI 82 : : : : : : : : :
QY	62 S-MGAKPTGSA--AANYTTAVDRPNPAYNKKHLHDAEFNTNAGFIALNIWDREDFVCTLGA 118 : : : : : : : : :
DB	83 SGMAAAPTAASGTSNTTVAADRSNFAYGKHLDQAEWCTNAAYLALNIWDREDFVCTLGA 142 : : : : : : : : :
QY	119 SNGVIRNSTAFNLVLGFYGKVTIVANELPNVLSNGVVVELTYDTSFWSVGCARGALWE 178 : : : : : : : : :
DB	143 SNGYFKASSDAFNVLGLIGLAGTDFA-NQRPNVEISOGIVELYTDYAFWSVGARGALWE 201 : : : : : : : : :
QY	179 CGCATLGAEEFOYAQSKEPKVEELNVICNVSOFYNKPKGYGVA--PFLPTDAGVATATGT 236 : : : : : : : : :
DB	202 CGCATLGAEEFOYAQSNPKIEMLNVTSSPAQFMHKHPGYKGTAAFPPLPVAAGTATATDT 261 : : : : : : : : :
QY	237 KSATINTHEMQVGASLSYRLNSLVPIYIGVOWSRATEDADNIRIAQPKLPVALNLTAWNP 296 : : : : : : : : :
DB	262 KSATVKYHEMQVGLSYRLNMLVPYIGVNWRATEDADTIRIAQPKLASAILNLJTWNP 321 : : : : : : : : :
QY	297 SLLENATALSTWDSFSDFPMQIVSCQINKFSRKACGVTVGATLVADADKSLTAEAKLINE 356 : : : : : : : : :
DB	322 TLLGVATTLDTSSNKYADFMOIVSMQINKMSRKACGIAGVATLTADLKWSITGEARHLIDE 381 : : : : : : : : :
QY	357 RAAHVSGOFRF 367 :
DB	382 RAAHINAQFRF 392 :
RESULT 6	
OMLE_CHLPS	STANDARD; PRT; 402 AA.
ID	OMLE_CHLPS STANDARD; PRT; 402 AA.
AC	P10332;
DT	01-MAR-1989 (Rel. 10, Created)
DT	01-MAR-1989 (Rel. 10, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
GN	OMPA OR OMP1.
OS	Chlamydia psittaci (Chlamydophila psittaci).
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX	NCBI_TaxID=83554;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=EAE A22/W;
RA	Pickett M.A., Everson S.J., Clarke I.N.;
RT	"Chlamydia psittaci ewe abortion agent: complete nucleotide sequence
RT	of the major outer membrane protein gene."
RL	FEMS Microbiol. Lett. 55:229-234(1988).
CC	-!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC	BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC	THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC	-!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC	MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC	-!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC	-----
CC	THIS SWISS-PROT entry is copyright. It is produced through a collabora-
CC	tion between the Swiss Institute of Bioinformatics and the EMBL Outstat-


```
SQ SEQUENCE 387 AA; 42009 MW; 4FD6FDC23248E0A2 CRC64;
Query Match 64.6%; Score 1358.5; DB 1; Length 387;
Best Local Similarity 66.7%; Pred. No. 6.8e-105;
Matches 246; Conservative 50; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEGAAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPLMIDGILWEGFGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDVKNQF 82
QY 62 SMGAKPTGSAANYTTA---VDRENPAYNKHLHDAEFTNAGFIALNWDVDFVCTGLG 118
DB 83 EMGAAPTGD--ADLTAPTPASRENPAYKHKMDAEFTNAAYNMALNWDVDFVCTGLG 140
QY 119 SNGYIRGNSTAFNLVGLFGVGTNNANLPSVLSNGVVELYTDTSFWSVSGARGALWE 178
DB 141 TSGYLKGNAAFNLVGLFGRETAVAADDIPNVLSQAVVELYTDTPAFWSVGARGAALWE 200
QY 179 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAPPLPTDAGVATATGKS 238
DB 201 CGCATLGAEFQYAOQSKPKVEELNVCNVAEFTINKPKYVGOEPLNKAQGVTSATDKD 260
QY 239 ATINYHEQVQASLSYRLNSLVPIGVOWSRATFDADNIRIAQKLPKPTAVNLNLTAWNPSL 298
DB 261 ASIDYHEQVQASLSYRLNSLVPIGVOWSRATFDADNIRIAQKLPKPTAVNLNLTAWNPSL 320
QY 299 LGNATALSTTDSFDMOIVSCQINKFKSRKACGVTVGATLVADKWSLTAFAERLINERA 358
DB 321 SSGSIDVT--KIDTTLQIVSLQNLKMKSRKSCGLAIGTTIIVADKYAVTVETRLIDERA 378
QY 359 AHVSGQFR 367
DB 379 AHVNAQFR 387

RESULT 8
OMIN_CHLTR STANDARD; PRT; 397 AA.
AC P13467;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H PRECURSOR (MOMP).
GN OMPA OR OMP13.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SEROVAR H;
RX MEDLINE=90045958; PubMed=2813066;
RA Hamilton P.T., Malinowski D.P.;
RT "Nucleotide sequence of the major outer membrane protein gene from Chlamydia trachomatis serovar H.;"
RL Nucleic Acids Res. 17:8366-8366(1989).
[2]
RN SEQUENCE FROM N.A.
RC STRAIN=H/UW-4;
RX Dean D.A.;
RT "Sequence analysis of the major outer membrane protein gene (ompA) of Chlamydia trachomatis.;"
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -
```

```
the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
DR EMBL; X16007; CA34145.1; -.
DR EMBL; AF304857; AAG41415.1; -.
DR PIR; S06589; MNCWTH; Chlamydia_OMP.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H.
SQ SEQUENCE 397 AA; 42946 MW; 478ACE3808BF37BA CRC64;

Query Match 63.7%; Score 1340.5; DB 1; Length 397;
Best Local Similarity 65.1%; Pred. No. 2.2e-103;
Matches 244; Conservative 51; Mismatches 71; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLIDGTIWEGAAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPLMIDGILWEGFGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDVKNKEF 82
QY 62 SMGAKPTGSAANY---TTAVDRPNPAYNKHLHDAEFTNAGFIALNWDVDFVCTGLG 117
DB 83 OMGAAPTNDAAIDQNDPKTNVARNPAYKHKMDAEFTNAAYNMALNWDVDFVCTGLG 142
QY 118 ASNGYIRGNSTAFNLVGLFGVGTNNANL---LPNYSLSNGVVELYTDTSFWSVSGARG 174
DB 143 ATTGYLKGNASFNVLVGLFGCTKSSDENTAKLVPTNALNRAVVELYTDTSFWSVSGARA 202
QY 175 ALWECGATLGAEFQYAOQSKPKVEELNVCNVSQFVNKPKYKGVAPPLPTDAGVATAT 234
DB 203 ALWECGATLGAEFQYAOQSKPKVEELNVCNVAEFTINKPKYVGOEPLDITAGTEAT 262
QY 235 GTSKATINYHEQVQASLSYRLNSLVPIGVOWSRATFDADNIRIAQKLPKPTAVNLNLTAW 294
DB 263 GTKDASIDYHEQVQASLSYRLNSLVPIGVOWSRATFDADNIRIAQKLPKAEILDVPTL 322
QY 295 NPSLLGNAT--ALSTTDSFDMOIVSCQINKFKSRKACGVTVGATLVADKWSLTAFAER 352
DB 323 NPTIAGKGTVVASGSDNDLADTMQIVSLQNLKMKSRKSCGLAIGTTIIVADKYAVTVETR 382
QY 353 LINERAAHVSGQFR 367
DB 383 LIDERRAAHVNAQFR 397

RESULT 9
OMIN_CHLTR STANDARD; PRT; 397 AA.
AC P23114;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3 PRECURSOR (MOMP).
GN OMPA OR OMP13.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=404 / SEROVAR L3;
RX MEDLINE=91285439; PubMed=2060793;
RA Fielder T.J., Peterson E.M., de la Maza L.M.;
RT "Nucleotide sequence of DNA encoding the major outer membrane protein of Chlamydia trachomatis serovar L3.;"
RL Gene 101:159-160(1991).
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
```

CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; X55700; CA39226.1; -
CC PIR; JE0413; JE0413.
CC InterPro; IPR000604; Chlamydia_OMP.
CC Pfam; PF01308; Chlamydia_OMP; 1.
CC ProDom; PD001717; Chlamydia_OMP; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 22 BY SIMILARITY.
CC CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3.
CC SEQUENCE 397 AA; 42885 MW; FIDDCF09535C2595 CRC64;

DR EMBL; X55700; CA39226.1; -
DR PIR; JE0413; JE0413.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
DR Outer membrane; Transmembrane; Porin; Signal.
DR SIGNAL 1 22 BY SIMILARITY.
DR CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3.
DR SEQUENCE 397 AA; 42885 MW; FIDDCF09535C2595 CRC64;

Query Match 63.6%; Score 1338.5; DB 1; Length 397;
Best Local Similarity 65.1%; Pred. No. 3.1e-103;
Matches 244; Conservative 49; Mismatches 73; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSLMIDGILWEGFGDPCDCTTWCDAISMRVGYGDFVDFRILKVDVNFKEF 82
QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKLHDAEFTNAGFTALNIWDRFVFCFLG 117
DB 83 QMGAEPTSDTAGLSNDPTTNAVRENPAYGKHMQDAEFTNAYMALNIWDRFVFCFLG 142
QY 118 ASNGYIRGNSTAFNLVGLFGVKTGVNANE---LPNLSNGVVELYTDTSFWSVSGARG 174
DB 143 ATTGYLKGNSASFNLVGLFGTKOSTNFNTAKLVPTNTALNQAVVELYTDITTFANVSGARA 202
QY 175 ALWECGCATLGAEOYQASQKPKVEELNVCNVSQFVSNKPKYGVAPPLPTDAGVATAT 234
DB 203 ALWECGCATLGAEOYQASQKPKVEELNVCNVSQFVSNKPKYGVAPPLPTDAGVATAT 262
QY 235 GTSKATINHEWQVAGSLSYRLNSLVPIYGVQWRSRATFDADNIRIAQPKLPTAVLNLTAW 294
DB 263 GTKDASIDYHEWQVAGSLSYRLNSLVPIYGVQWRSRATFDADNIRIAQPKLPTAVLNLTAW 322
QY 295 NPSLL--GNATALSTDSFDFMQIVSQINKFKSRKACGVTVGATLVADKWSLTAPAR 352
DB 323 NPTTAGKGVVAGSGSENELADTMQIVSLQNLNKMKSRSKSGIAGVTIVDADKYAVTVETR 382
QY 353 LINERAAHVSGQFRF 367
DB 383 LIDERAAHVNAQFRF 397

RESULT 10
OM1A_CHLTR STANDARD; PRT; 396 AA.
AC P23732;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A PRECURSOR (MOMP).
GN OMPA OR OMPLA.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Sal/OT / SEROVAR A;

RX MEDLINE=91045088; PubMed=2235504;
RA Hayes L.J., Clarke I.N.;
RT "Nucleotide sequence of the major outer membrane protein gene of
Chlamydia trachomatis strain A/Sal/OT";
RL Nucleic Acids Res. 18:6136-6136(1990).
CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M58938; AAA23141.1; -
CC EMBL; M33635; AAA92785.1; -
CC PIR; SI2799; SI2799.
CC InterPro; IPR000604; Chlamydia_OMP.
CC Pfam; PF01308; Chlamydia_OMP; 1.
CC ProDom; PD001717; Chlamydia_OMP; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 22
CC CHAIN 23 396 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A.
CC SEQUENCE 396 AA; 42877 MW; 2F9D3B0CE2D08162 CRC64;

Query Match 63.0%; Score 1325; DB 1; Length 396;
Best Local Similarity 64.7%; Pred. No. 4.1e-102;
Matches 242; Conservative 47; Mismatches 77; Indels 8; Gaps 3;

QY 2 LPVGNPSPSLLIDGTIWEAGAAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSLMIDGILWEGFGDPCDCTTWCDAISMRVGYGDFVDFRILKVDVNFKEF 82
QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKLHDAEFTNAGFTALNIWDRFVFCFLG 117
DB 83 QMGAEPTSDTAGLSNDPTTNAVRENPAYGKHMQDAEFTNAYMALNIWDRFVFCFLG 142
QY 118 ASNGYIRGNSTAFNLVGLFGVKTGVNANE---LPNLSNGVVELYTDTSFWSVSGARG 174
DB 143 ATTGYLKGNSASFNLVGLFGTKOSTNFNTAKLVPTNTALNQAVVELYTDITTFANVSGARA 202
QY 175 ALWECGCATLGAEOYQASQKPKVEELNVCNVSQFVSNKPKYGVAPPLPTDAGVATAT 234
DB 203 ALWECGCATLGAEOYQASQKPKVEELNVCNVSQFVSNKPKYGVAPPLPTDAGVATAT 262
QY 235 GTSKATINHEWQVAGSLSYRLNSLVPIYGVQWRSRATFDADNIRIAQPKLPTAVLNLTAW 294
DB 263 GTKDASIDYHEWQVAGSLSYRLNSLVPIYGVQWRSRATFDADNIRIAQPKLPTAVLNLTAW 322
QY 295 NPSLLGNATALSTDS--FDFMQIVSQINKFKSRKACGVTVGATLVADKWSLTAEARL 353
DB 323 NPTTAGKGVVAGSGSENELADTMQIVSLQNLNKMKSRSKSGIAGVTIVDADKYAVTVETR 382
QY 354 LINERAAHVSGQFRF 367
DB 383 LIDERAAHVNAQFRF 396

RESULT 11
OM1E_CHLTR STANDARD; PRT; 393 AA.
AC P17451;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR E PRECURSOR (MOMP).

GN OMPA OR OMPLE.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=813;

RP SEQUENCE FROM N.A.

RC STRAIN=BOUR / SEROVAR E;

EX MEDLINE=90287737; PubMed=2356137;

RA Peterson E.M., Markoff B.A., de la Maza L.M.;

RT "The major outer membrane protein nucleotide sequence of Chlamydia

trachomatis, serovar E.;"

RL Nucleic Acids Res. 18:3414-3414(1990).

CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY

BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH

CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.

CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP

CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.

CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; X52557; CAA36791.1;

DR PIR; S10201; MMCWTE.

DR InterPro; IPR000604; Chlamydia_OMP.

DR Pfam; PF01308; Chlamydia_OMP; 1.

DR ProDom; PD001717; Chlamydia_OMP; 1.

DR Outer membrane; Transmembrane; Porin; Signal.

FT CHAIN

FT SIGNAL

FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR E.

SQ SEQUENCE 393 AA; 42424 MW; AB2B82D16027B361 CRC64;

Query Match 62.6%; Score 1315.5; DB 1; Length 393;

Best Local Similarity 65.3%; Pred. No. 2.5e-101;

Matches 243; Conservative 46; Mismatches 76; Indels 7; Gaps 4;

Qy 2 LPVGNPSPDLLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61

Db 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISRMVGYGDFVDFRILKTDVNFKEF 82

Qy 62 SMGAKP---TGSAAAYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFDFVCTLGA 118

Db 83 QMGDKPTSTTGNATAP-TTLTARENPAAYGRHMQDAEMFTNAACMALNIWDRFDFVCTLGA 141

Qy 119 SNGYIRGNSTAFNLVGLFG--VKGTNNANLNPVSLNSGVVELYTDTSFSWSVGARGAL 176

Db 142 SSGYLKGNASFNVLGFLGDENENSTVKNTPVNSLDQSVVELYTDTSFSWSVGARGAL 201

Qy 177 WECCGATLGAFQYQAQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGT 236

Db 202 WECCGATLGAFQYQAQSKPKVEELNVCNVAEFTNPKGYGVQEFPLALITAGTATGT 261

Qy 237 KSATINHEWQVQASLSYRLNSLPYIGVQWSRATFDADNTRIAQPKLPTAVLNLTAWN 296

Db 262 KDAIDYHEWQASLSYRLNFTPIYIGVQWSRASFDADTTRIAQPKSATIFDTTINP 321

Qy 297 SLILGNATALSTWD-SFSDMOIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLIN 355

Db 322 TIAGAGDVKAAGQLGDMIVSLQLNKMKSRSKSGIAGVTITIVDADKAVYVETRLID 381

Qy 356 ERAAHVSGQFRF 367

Db 382 ERAAHVNAQFRF 393

RESULT 12

OMIC_CHLTR

ID OMIC_CHLTR

AC P08780; STANDARD;

DT 01-NOV-1988 (Rel. 09, Created)

DT 01-NOV-1988 (Rel. 09, Last sequence update)

DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR C PRECURSOR (MOMP).

GN OMPA OR OMP1 OR OMP1C.

OS Chlamydia trachomatis.

OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

OX NCBI_TaxID=813;

RP SEQUENCE FROM N.A.

EX MEDLINE=87307955; PubMed=3040664;

RA Stephens R.S., Sanchez-Pescador R., Wagar E.A., Inouye C., Urdea M.S.;

RT "Diversity of Chlamydia trachomatis major outer membrane protein

genes.;"

RL J. Bacteriol. 169:3879-3885(1987).

RN [2]

RP SEQUENCE FROM N.A.

RC STRAIN=C/TW3;

EX MEDLINE=20407420; PubMed=10950788;

RA Dean D., Suchland R.J., Stamm W.E.;

RT "Evidence for long-term cervical persistence of Chlamydia trachomatis

by omp1 genotyping.;"

RL J. Infect. Dis. 182:909-916(2000).

CC -1- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY

BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH

CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.

CC -1- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP

CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.

CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.

CC -1- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.

CC -----

CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

CC -----

DR EMBL; M17343; AAA23156.1;

DR EMBL; AF202455; AAG09443.1;

DR PIR; S11011; MMCWTC.

DR InterPro; IPR000604; Chlamydia_OMP.

DR Pfam; PF01308; Chlamydia_OMP; 1.

DR ProDom; PD001717; Chlamydia_OMP; 1.

DR Outer membrane; Transmembrane; Porin; Signal.

FT CHAIN

FT SIGNAL

FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR C.

SQ SEQUENCE 397 AA; 42892 MW; 0047BCDB108E5309 CRC64;

Query Match 62.6%; Score 1315.5; DB 1; Length 397;

Best Local Similarity 64.3%; Pred. No. 2.5e-101;

Matches 241; Conservative 49; Mismatches 76; Indels 9; Gaps 3;

Qy 2 LPVGNPSPDLLIDGTWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61

Db 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISRMVGYGDFVDFRILKTDVNFKEF 82

Qy 62 SMGAKP---TGSAAAYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFDFVCTLGA 117

Db 83 QMGARPTSDVAGLQNDPTINVARPNPAYKHMQDAEMFTNAAYMALNIWDRFDFVCTLGA 142

Qy 118 ASNGYIRGNSTAFNLVGLFGYKGTNNANLNPVSLNSGVVELYTDTSFSWSVGARG 174

Db 143 ATTGYLKNASFNVLGFLGFTKQSSSFNTAKLIPNTALNEAVVELYINTTFAWSVGARA 202

Qy 175 ALWECGATLGAFQYQAQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATAT 234

```
Db 203 ALWECGCATLGASFOYAQSKPKVEELNVLNASEFTINKPKYVGAEPFLNITAGTEAAT 262
Qy 235 GTSKATINHEWQVQASLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294
Db 263 GTKASIDYHEWQASLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 322
Qy 295 NPSLLGNATLALS--TTDSFDMQIVSCQINKFKSKACGVTVGATLVADKWSLTAEAR 352
Db 323 NRTTAGKSVYSAGTDELADMTQIVSLQNLNKKSRKSCGIAVGTGTTIVDAKYAVTVEAR 382
Qy 353 LINERAAHVSGQFRF 367
Db 383 LIDERRAAHVNAQFRF 397

RESULT 13
OM1L_CHLTR
ID OM1L_CHLTR STANDARD; PRT; 393 AA.
AC P19542;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1 PRECURSOR (MOMP).
GN OMPA OR OMP1L.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RA Pickett M.A., Ward M.E., Clarke I.N.;
RT "Complete nucleotide sequence of the major outer membrane protein
gene from Chlamydia trachomatis serovar L1.";
RL FEMS Microbiol. Lett. 42:185-190(1987).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
CC EMBL; M36533; AAA23142.1; -.
CC PIR; S06259; S06259.
CC InterPro; IPR000604; Chlamydia_OMP.
CC Pfam; PF01308; Chlamydia_OMP; 1.
CC ProDom; PD001171; Chlamydia_OMP; 1.
CC Outer membrane; Transmembrane; Porin; Signal.
CC SIGNAL 1 22
FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1.
SQ SEQUENCE 393 AA; 42543 MW; 7A952839408EE2DF CRC64;
```

Query Match 62.58; Score 1313.5; DB 1; Length 393;
Best Local Similarity 64.58; Pred. No. 3.6e-101;
Matches 240; Conservative 51; Mismatches 74; Indels 7; Gaps 4;

```
Qy 2 LPVGNPSPSLIDGTIWEAGAAGPCDPCATWCDAISLRAGFYGDYFDRILKYDPAKTF 61
Db 23 LPVGNPAEPLMIDGILLWEGFGDPCDCTTWCDAISNRWGYGDFVDRVLQTDVNEKF 82
Qy 62 SMGAKP---TGSAAANYTAVDRPNPAYNKLHDAENFTNAGFTALNIWDRFDVFTLGA 118
Db 83 QMGAKPTATTGNAAP-STCTARENPAYGRHMQDAENFTNAAAYMALNIWDRFDVFTLGA 141
Qy 119 SNGYIRGNSTAFNLVGLFG--VKGTTVNAELPNVLSLNGVVELYTDTSFSWSVGARGAL 176
```

```
Db 142 TSGLKNGSASFNLVGLFGDNEQSTVKKDAVPNMSFDQSVVELYTDTSFWSVGARGAL 201
Qy 177 WECCATILGAEFOYAQSKPKVEELNVLNASEFTINKPKYVGAEPFLNITAGTEAAT 236
Db 202 WECCATILGASFOYAQSKPKVEELNVLNASEFTINKPKYVGAEPFLNITAGTEAAT 261
Qy 237 KEATINHEWQVQASLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 296
Db 262 KASIDYHEWQASLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 321
Qy 297 SLLGNATLSTTD--SFSDFMQIVSCQINKFKSKACGVTVGATLVADKWSLTAEARLN 355
Db 322 TIAGAGEVKANAEGQLGDTMQIVSLQNLNKKSRKSCGIAVGTGTTIVDAKYAVTVETRLD 381
Qy 356 BRAAHVSGQFRF 367
Db 382 ERAAHVNAQFRF 393
```

RESULT 14
OM1D_CHLTR
ID OM1D_CHLTR STANDARD; PRT; 393 AA.
AC Q46409;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D PRECURSOR (MOMP).
GN OMPA OR OMP1 OR CT681.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-D/B-120;
RC MEDLINE=93013014; PubMed=1398119;
RX Stothard D.R., Boguslawski G., Jones R.B.;
RT "Phylogenetic analysis of the Chlamydia trachomatis major outer
membrane protein and examination of potential pathogenic
determinants.";
RL Infect. Immun. 66:3618-3625(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-D/IU-71960;
RX MEDLINE=98339860; PubMed=9673241;
RA Stothard D.R., Boguslawski G., Jones R.B.;
RT "Phylogenetic analysis of the Chlamydia trachomatis major outer
membrane protein and examination of potential pathogenic
determinants.";
RL Infect. Immun. 66:3618-3625(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-D/UW-3/CX;
RX MEDLINE=9900809; PubMed=9784136;
RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
RA Davis R.W.;
RT "Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis.";
RL Science 282:754-759(1998).

-!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/

or send an email to license@isb-sib.ch.

CC EMBL; X62918; CAA44701.1; -
 CC EMBL; AF063195; AAC31436.2; -
 CC EMBL; AE001338; AAC68276.1; -
 CC InterPro; IPR000604; Chlamydia_OMP.
 CC Pfam; PF01308; Chlamydia_OMP; 1.
 CC ProDom; PD001717; Chlamydia_OMP; 1.
 CC Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
 KW SIGNAL 1 22 BY SIMILARITY.
 FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D.
 SQ SEQUENCE 393 AA; 42438 MW; 8CD692FD3EEF21D6 CRC64;

Query Match 62.2%; Score 1308.5; DB 1; Length 393;
 Best Local Similarity 64.8%; Pred. No. 9.3e-101;
 Matches 241; Conservative 48; Mismatches 76; Indels 7; Gaps 4;
 QY 2 LPVGNPDPSSLIDGTIWEAGAAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
 DB 23 LPVGNPAEPSLMDIGILWEGFGDPCPCATWCDAISMRVGYGDFVDRVLKTDVKNKEF 82
 QY 62 SMGAKP---TGSAANTTAVDRPNPAYNKHLDHDAEFTNAGFTIALNIWDRFDFVCTLGA 118
 DB 83 QMGAKPTTDTGNSAAP-STLTARENPAYGRHMQDAEFTNAAACMALNIWDRFDFVCTLGA 141
 QY 119 SNGYIRGNSTAFNLVGLFG--VKGTTYNANELPNVSLNSGVVELYTDTSFWSVGARGAL 176
 DB 142 TSGYLKGNASAFNLVGLFGDQENQKTVKAEVPMNSFDQSVVELYTDTSFWSVGARGAL 201
 QY 177 WECGCATLGAEFOYAQSKPKVEELNVCNVSQFSVKNPKGYKGVAFLPTDAGVATATGT 236
 DB 202 WECGCATLGASFOYAQSKPKVEELNVCNVAEFTINKPKGVGKEFFPLDLTAGTDAATGT 261
 QY 237 KSATINHEWQVGSASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 296
 DB 262 KDAIDYHEWQASLSYRLNMFPTPIGVKWSRASFDADTIRIAQPKSATAIFDTTLNP 321
 QY 297 SLGNATALSTTD--SFSDFMQIVSCQINKFKSKACGVTVGATLVADKWSLTAEARLI 355
 DB 322 TIAGGVKVTGAEGQLGDTMQIVSLQNLKMKSRKSCGIAGVTIVDADKAVIVETRLID 381
 QY 356 ERAAHVSGQFRF 367
 DB 382 ERAAHVNAQFRF 393

RESULT 15
 OMIF_CHLTR STANDARD; PRT; 395 AA.
 AC P16155;
 DT 01-APR-1990 (Rel. 14, Created)
 DT 01-APR-1990 (Rel. 14, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR F PRECURSOR (MOMP).
 GN OMPA OR OMPIF.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=IC-CAL3 / SEROVAR F;
 RX MEDLINE=90192102; PubMed=2315025;
 RA Zhang Y.X., Morrison S.G., Caldwell H.D.;
 RT "The nucleotide sequence of major outer membrane protein gene of
 Chlamydia trachomatis serovar F.";
 RL Nucleic Acids Res. 18:1061-1061(1990).
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.

CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC EMBL; X52080; CAA36299.1; -
 CC PIR; S08463; MMCWTF.
 CC InterPro; IPR000604; Chlamydia_OMP.
 CC Pfam; PF01308; Chlamydia_OMP; 1.
 CC ProDom; PD001717; Chlamydia_OMP; 1.
 KW Outer membrane; Transmembrane; Porin; Signal.
 FT SIGNAL 1 22
 FT CHAIN 23 395 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR F.
 SQ SEQUENCE 395 AA; 42586 MW; 7F90FFDEEC264ACF CRC64;

Query Match 62.2%; Score 1307.5; DB 1; Length 395;
 Best Local Similarity 63.5%; Pred. No. 1.1e-100;
 Matches 237; Conservative 53; Mismatches 76; Indels 7; Gaps 3;

QY 2 LPVGNPDPSSLIDGTIWEAGAAGDPCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
 DB 23 LPVGNPAEPSLMDIGILWEGFGDPCPCATWCDAISMRVGYGDFVDRVLKTDVKNKEF 82
 QY 62 SMG---AKPTGSAANTTAVDRPNPAYNKHLDHDAEFTNAGFTIALNIWDRFDFVCTLGA 118
 DB 83 EMGEALAGASGNTTSLTKLVERTNPAYGKHMQDAEFTNAACTNLNIWDRFDFVCTLGA 142
 QY 119 SNGYIRGNSTAFNLVGLF--GVKGTTYNANELPNVSLNSGVVELYTDTSFWSVGARGAL 176
 DB 143 TSGYLKGNASAFNLVGLFGDQENQKTVKAEVPMNSFDQSVVELYTDTSFWSVGARGAL 202
 QY 177 WECGCATLGAEFOYAQSKPKVEELNVCNVSQFSVKNPKGYKGVAFLPTDAGVATATGT 236
 DB 203 WECGCATLGASFOYAQSKPKVEELNVCNVAEFTINKPKGVGKEFFPLDLTAGTDAATGT 262
 QY 237 KSATINHEWQVGSASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 296
 DB 263 KDAIDYHEWQASLSYRLNMFPTPIGVKWSRASFDADTIRIAQPKRLVTPVVDITLNP 322
 QY 297 SL--LGNATALSTTD--SFSDFMQIVSCQINKFKSKACGVTVGATLVADKWSLTAEARLI 354
 DB 323 TIACGSGVAGANTEGQISDTMQIVSLQNLKMKSRKSCGIAGVTIVDADKAVIVETRLI 382
 QY 355 NERAAHVSGQFRF 367
 DB 383 DERAHVNAQFRF 395

Search completed: February 7, 2002, 21:42:47
 Job time: 601 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:19 ; Search time 172 Seconds
(without alignments)
335.066 Million cell updates/sec

Title: US-09-391-606-15

Perfect score: 2103

Sequence: 1 MLPVGNPSPDLLIDTWE.....OKLISEEDLNSAVDHHHHH 394

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL17:**
- 2: sp_archaea:**
- 3: sp_bacteria:**
- 4: sp_fungi:**
- 5: sp_human:**
- 6: sp_invertebrate:**
- 7: sp_mhc:**
- 8: sp_organelle:**
- 9: sp_phage:**
- 10: sp_plant:**
- 11: sp_rodent:**
- 12: sp_virus:**
- 13: sp_vertebrate:**
- 14: sp_unclassified:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1910	90.8	389	Q08085	Q08085 chlamydia p
2	1531.5	72.8	389	Q9APM4	Q9APM4 chlamydophi
3	1529.5	72.7	388	Q9AIK1	Q9AIK1 chlamydia p
4	1501.5	71.4	391	Q46235	Q46235 chlamydia p
5	1497.5	71.2	388	Q9AIK0	Q9AIK0 chlamydia p
6	1497.5	71.2	389	Q9AIH9	Q9AIH9 chlamydophi
7	1487	70.7	392	Q9AIJ4	Q9AIJ4 chlamydia p
8	1480	70.4	381	Q9AIJ2	Q9AIJ2 chlamydophi
9	1479	70.3	390	Q9AIJ5	Q9AIJ5 chlamydia p
10	1477	70.2	392	Q99QB0	Q99QB0 chlamydophi
11	1469.5	69.9	380	Q9AIJ1	Q9AIJ1 chlamydophi
12	1423	67.7	391	Q9AIJ2	Q9AIJ2 chlamydia p
13	1420.5	67.5	341	Q9X717	Q9X717 chlamydophi
14	1419	67.5	402	Q46193	Q46193 chlamydia p
15	1415	67.3	382	Q9AIJ9	Q9AIJ9 chlamydia p
16	1415	67.3	395	Q9AIJ7	Q9AIJ7 chlamydia p
17	1415	67.3	397	Q9AIJ8	Q9AIJ8 chlamydia p
18	1413	67.2	402	Q9AIJ6	Q9AIJ6 chlamydia p
19	1410	67.0	402	Q9AIJ0	Q9AIJ0 chlamydia p

20	1408	67.0	402	2	Q46203	Q46203 chlamydia p
21	1407	66.9	402	2	Q46236	Q46236 chlamydia p
22	1382	65.7	402	2	Q9AIJ3	Q9AIJ3 chlamydia p
23	1371.5	65.2	352	2	Q70085	Q70085 chlamydia p
24	1367.5	65.0	352	2	Q70050	Q70050 chlamydia p
25	1363.5	64.8	352	2	Q69307	Q69307 chlamydia p
26	1362.5	64.8	352	2	Q69306	Q69306 chlamydia p
27	1357	64.5	337	2	Q9XBF6	Q9XBF6 chlamydophi
28	1353	64.3	336	2	Q9XBF5	Q9XBF5 chlamydophi
29	1353	64.3	353	2	Q69305	Q69305 chlamydia p
30	1352	64.3	396	2	Q9AIJ8	Q9AIJ8 chlamydia s
31	1351	64.2	404	2	Q46407	Q46407 chlamydia t
32	1343	63.9	389	2	Q9AIJ4	Q9AIJ4 chlamydia s
33	1342	63.8	386	2	Q69093	Q69093 chlamydia t
34	1341.5	63.8	386	2	Q9AIJ5	Q9AIJ5 chlamydia s
35	1339.5	63.7	397	2	Q69094	Q69094 chlamydia t
36	1337.5	63.6	397	2	Q69095	Q69095 chlamydia t
37	1336.5	63.6	387	2	Q9AIJ0	Q9AIJ0 chlamydia s
38	1335.5	63.5	397	2	Q9F951	Q9F951 chlamydia t
39	1333	63.4	340	2	Q9XBF2	Q9XBF2 chlamydophi
40	1331.5	63.3	387	2	Q9AIJ1	Q9AIJ1 chlamydia s
41	1331.5	63.3	397	2	Q9F950	Q9F950 chlamydia t
42	1330.5	63.3	385	2	Q9AIJ6	Q9AIJ6 chlamydia s
43	1328	63.1	396	2	Q46406	Q46406 chlamydia t
44	1327.5	63.1	385	2	Q9AIJ7	Q9AIJ7 chlamydia s
45	1326	63.1	356	2	Q52924	Q52924 chlamydia p

ALIGNMENTS

RESULT 1

Q08085 PRELIMINARY; PRT; 389 AA.
AC Q08085;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
OS Chlamydia psittaci (Chlamydophila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOALA TYPE 1;
RX MEDLINE=94171025; PubMed=8125292;
RA Girjes A.A., Carrick F.N., Lavin M.F.;
RT "Remarkable sequence relatedness in the DNA encoding the major outer
RT membrane protein of Chlamydia psittaci (koala type 1) and Chlamydia
RT pneumoniae.";
RL Gene 138:139-142(1994).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
DR EMBL; J2023; CAA50906.1;
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR PRINTS; PR01334; CHLAMIDIACMP.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41579 MW; 5DC50E85A6F4E50F CRC64;

Query Match 90.8%; Score 1910; DB 2; Length 389;
Best Local Similarity 97.5%; Pred. No. 2.1e-146;
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 2 LPVGNPSPDLLIDTWEAGAGPCDPCATWCDAISLRAGFYGDYVDFRLKVDAPKTF 61

```

Db 24 LPVGNPSDPSLLIDGTIWEAGAAGDPCDPCATWCDAISLRAGFYGYDFDRILKVDAPKTF 83
QY 62 SMGAKPTGSAANNTTAVDRPNPAINKHLHDAEFTNAGFIALNIWDRFDFCTLGASNG 121
Db 84 SMGAKPTGSAANTTAVDRPNPAINKHLHDAEFTNAGFIALNIWDRFDFCTLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 181
Db 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 203
QY 182 ATLAGAEFYAQSKPKVEELNVICNVQSVNPKPKYGVAFPLPTDAGVATATGKTSATI 241
Db 204 ATLAGAEFYAQSKPKVEELNVICNVQSVNPKPKYGVAFPLPTDAGVATATGKTSATI 263
QY 242 NYHEWQVGLASLYRLNSLVPIGVQWSRATFDADNIRIAQPKLPPTAVLNLTAWNPSLLGN 301
Db 264 NYHEWQVGLASLYRLNSLVPIGVQWSRATFDADNIRIAQPKLPPTAVLNLTAWNPSLLGN 323
QY 302 ATALSTTDSFDFMQIYSCQINKFKSRKACGTVTGATLVADADKWSLTAEARLINERAHV 361
Db 324 TTTATSDSFDFMQIYSCQINKFKSRKACGTVTGATLVADADKWSLTAEARLINERAHV 383
QY 362 SQGFRF 367
Db 384 SQGFRF 389

```

RESULT 2

```

Q9APM4
ID Q9APM4 PRELIMINARY; PRT; 389 AA.
AC Q9APM4;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMP1.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LLG;
RX MEDLINE=20569239; PubMed=11119563;
RA Vreton E., Psarrou E., Kaisar M., Vilisidou I., Salti-Montesanto V.,
Longbottom D.;
RT "Identification of protective epitopes by sequencing of the major
outer membrane protein gene of a variant strain of Chlamydia psittaci
serotype 1."
RL Infect. Immun. 69:607-612(2001).
DR EMBL; AF272945; AAG53881.1; -.
KW Signal.
FT SIGNAL.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41897 MW; 2051369C7DBAAF5 CRC64;

```

```

Query Match 72.8%; Score 1531.5; DB 2; Length 389;
Best Local Similarity 75.7%; Pred. No. 8.2e-116;
Matches 278; Conservative 38; Mismatches 50; Indels 1; Gaps 1;

```

```

QY 2 LPVGNPSDPSLLIDGTIWEAGAAGDPCDPCATWCDAISLRAGFYGYDFDRILKVDAPKTF 61
Db 23 LPVGNPAEPSLLIDGTIWEAGASGDPDPCSTWCDAISIRAGYGYDFDRILKVDVNKTI 82
QY 62 S-MGAKPTGSAANNTTAVDRPNPAINKHLHDAEFTNAGFIALNIWDRFDFCTLGASN 120
Db 83 TGMGAVPTGTAADYKTPTRNPATAYGKHLQDAEFTNAAFLALNIWDRFDFCTLGASN 142
QY 121 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 180
Db 143 GYFKASSAFLNLVGLIGVKGSSAADQLPNVGIQTGIVEFTDTSFWSVGARGALWECG 202

```

```

QY 181 CATLAGAEFYAQSKPKVEELNVICNVQSVNPKPKYGVAFPLPTDAGVATATGKTSAT 240
Db 203 CATLAGAEFYAQSKPKVEELNVICNVQSVNPKPKYGVAFPLPTDAGVATATGKTSAT 262
QY 241 INYHEWQVGLASLYRLNSLVPIGVQWSRATFDADNIRIAQPKLPPTAVLNLTAWNPSLLG 300
Db 263 IKYHEWQVGLASLYRLNSLVPIGVQWSRATFDADNIRIAQPKLAAAVLNLTWNPTLLG 322
QY 301 NATALSTTDSFDFMQIYSCQINKFKSRKACGTVTGATLVADADKWSLTAEARLINERAHV 360
Db 323 EATTLDTNFKFADFLQIASIQINKMKSRKACGVAAGATLIDADKWSITGEARLINERAHV 382
QY 361 VSGQFRF 367
Db 383 MNAQFRF 389

```

RESULT 3

```

Q9AIK1
ID Q9AIK1 PRELIMINARY; PRT; 388 AA.
AC Q9AIK1;
DT 01-JUN-2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia phila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VS225;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae."
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269259; AAK0240.1; -.
KW Signal.
FT NON_TER 1 1 POTENTIAL.
FT SIGNAL <1 19 MAJOR OUTER MEMBRANE PROTEIN.
FT CHAIN 20 388
SQ SEQUENCE 388 AA; 41573 MW; 8E23D22C9B9948D CRC64;

```

```

Query Match 72.7%; Score 1529.5; DB 2; Length 388;
Best Local Similarity 76.2%; Pred. No. 1.2e-115;
Matches 281; Conservative 35; Mismatches 50; Indels 3; Gaps 2;

```

```

QY 2 LPVGNPSDPSLLIDGTIWEAGAAGDPCDPCATWCDAISLRAGFYGYDFDRILKVDAPKTF 61
Db 20 LPVGNPAEPSLLIDGTIWEAGASGDPDPCATWCDAISIRAGYGYDFDRILKVDVNKTI 79
QY 62 S-MGAKPTGSAANNTTAVDRPNPAINKHLHDAEFTNAGFIALNIWDRFDFCTLGASN 120
Db 80 SGMAAPTGSAADYKTPTRNPATAYGKHLQDAEFTNAAFLALNIWDRFDFCTLGASN 139
QY 121 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 180
Db 140 GYFKASSAFLNLVGLIGVKGSSAADQLPNVGIQTGIVEFTDTSFWSVGARGALWECG 199
QY 181 CATLAGAEFYAQSKPKVEELNVICNVQSVNPKPKYGVAFPLPTDAGVATATGKTS 238
Db 200 CATLAGAEFYAQSKPKVEELNVICNVQSVNPKPKYGVAFPLPTDAGVATATGKTS 259
QY 239 ATINYHEWQVGLASLYRLNSLVPIGVQWSRATFDADNIRIAQPKLPPTAVLNLTAWNPSL 298
Db 260 ATLKYHEWQVGLASLYRLNSLVPIGVQWSRATFDADNIRIAQPKLAAAVLNLTWNPTL 319
QY 299 LGNATALSTTDSFDFMQIYSCQINKFKSRKACGTVTGATLVADADKWSLTAEARLINERA 358
Db 320 LGNATALSTTDSFDFMQIYSCQINKFKSRKACGTVTGATLVADADKWSLTAEARLINERA 379
QY 359 AHVSGQFRF 367

```

```
Db 380 AHMNAQFRF 388
|||||
RESULT 4
Q46235
ID Q46235 PRELIMINARY; PRT; 391 AA.
AC Q46235;
DT 01-NOV-1996 (TremBLrel. 01, Created)
DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN MOMP OR OMPA.
OS Chlamydia psittaci (Chlamydophila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamyidophila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AVIAN TYPE C;
RA Storey C.C.;
RL Submitted (MAR-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-GERMAN DUCK;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; L25436; AAK0241.1; -.
DR EMBL; AF269261; AAK0242.1; -.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR PRINTS; PR01334; CHLAMIDIACOMP.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Signal.
FT SIGNAL.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 391 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 391 AA; 42427 MW; D98A8F5E76362A00 CRC64;

Query Match 71.4%; Score 1501.5; DB 2; Length 391;
Best Local Similarity 73.2%; Pred. No. 2.2e-113;
Matches 270; Conservative 48; Mismatches 48; Indels 3; Gaps 2;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDCAISLRAGFYGDYVDFRLKVDPAKTF 61
Db 23 LPVGNPAEPLSLIDGTIWEAGAGDPCDCAISLRAGFYGDYVDFRLKVDVNTF 82
QY 62 S-MGAKPTGSAANVTAVDRNPAYNKLHDAEFTNAGFTALNIWDRFVDFCTLGASN 120
Db 83 SGIGKKPTGSSPNDFKNAEDRPNVAYGRHLQDSEFTNAAFLALNIWDRFDFCTLGASN 142
QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLPNLSNGVVELYDTDSFSVSGARGALWECG 180
Db 143 GYFKASSAANFLVGLIGVKGSSLTNDQLPNVAITGVVEFTDTTFSVSGARGALWECG 202
QY 181 CATLGAEFQYAOQSKPVEELNVCNVSQFSVKNPKYKQVA--FPLPTDAGVATATGKS 238
Db 203 CATLGAEFQYAOQSNPKIEMNLNVISSPAQFVVHKPRGYKGTSAFPLPANAGTEAATDTKS 262
QY 239 ATINYHEQVGLASLYRLNSLPYIGVQVSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 298
Db 263 ATLYKHEQVGLASLYRLNMLVPYIGVWNSRATFDADTIRIAQPKLASAVMNLTTWNPTL 322
QY 299 LGNATALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINERA 358
Db 323 LGNATALSTDSFDMQIVSCQINKFKSRKACGLAIGNTLIDADKWSITGEARLINERA 382
QY 359 AHVSCQFRF 367
Db 383 AHMNAQFRF 391
|||||

Query Match 71.2%; Score 1497.5; DB 2; Length 388;
Best Local Similarity 72.9%; Pred. No. 4.6e-113;
Matches 269; Conservative 49; Mismatches 48; Indels 3; Gaps 2;

QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDCAISLRAGFYGDYVDFRLKVDPAKTF 61
Db 20 LPVGNPAEPLSLIDGTIWEAGAGDPCDCAISLRAGFYGDYVDFRLKVDVNTF 79
QY 62 S-MGAKPTGSAANVTAVDRNPAYNKLHDAEFTNAGFTALNIWDRFVDFCTLGASN 120
Db 80 SGIGKKPTGSSPNDFKNAEDRPNVAYGRHLQDSEFTNAAFLALNIWDRFDFCTLGASN 139
QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLPNLSNGVVELYDTDSFSVSGARGALWECG 180
Db 140 GYFKASSAANFLVGLIGVKGSSLTNDQLPNVAITGVVEFTDTTFSVSGARGALWECG 199
QY 181 CATLGAEFQYAOQSKPVEELNVCNVSQFSVKNPKYKQVA--FPLPTDAGVATATGKS 238
Db 200 CATLGAEFQYAOQSNPKIEMNLNVISSPAQFVVHKPRGYKGTSAFPLPANAGTEAATDTKS 259
QY 239 ATINYHEQVGLASLYRLNSLPYIGVQVSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 298
Db 250 ATLYKHEQVGLASLYRLNMLVPYIGVWNSRATFDADTIRIAQPKLASAVMNLTTWNPTL 319
QY 299 LGNATALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINERA 358
Db 320 LGNATALSTDSFDMQIVSCQINKFKSRKACGLAIGNTLIDADKWSITGEARLINERA 379
QY 359 AHVSCQFRF 367
Db 380 AHMNAQFRF 388
|||||

RESULT 6
Q9AIH9
ID Q9AIH9 PRELIMINARY; PRT; 389 AA.
AC Q9AIH9;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMPA.
OS Chlamydophila caviae.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamyidophila.
OX NCBI_TaxID=83557;
RN [1]
```

```

DR      EMBL; AF269267; AAK00248.1; -.
KW      Signal.
FT      SIGNAL.
FT      CHAIN 1 22 POTENTIAL.
FT      CHAIN 23 392 MAJOR OUTER MEMBRANE PROTEIN.
SQ      SEQUENCE 392 AA; 42293 MW; FC31FC051955246C CRC64;

      Query Match      70.7%; Score 1487; DB 2; Length 392;
      v Best Local Similarity 74.2%; Pred. No. 3.3e-112;
      : Matches 276; Conservative 39; Mismatches 49; Indels 8; Gaps
      QY      2 LPVGNPSDFSLIDGTINWGAAGDCPCDPCATWCDAISLRAGSYGYDVRILKVDAPKTF 61
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      db      23 LPVGNPAEPSLLIDGTINWEGASGDCPCDPCATWCDAISLRAGSYGYDVRVLKVDVNTKF 82
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

      QY      62 SMGAK ----PTGSAANYTAVDRNPAYNKKLHDAEFTNAGFTALNIWDRFDVFCITG 117
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
      db      83 SGMAKPTETATGASAT--TAVDRNLAYGKKHQDAEFTNAAFTALAIWDRFDICTCTLG 141
      |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

```

Query Match          : 70.7%; Score 1487; DB 2; Length 392;
Best Local Similarity : 74.2%; Pred. No. 3.3e-112;

```

[illegible]

RESULT	8
Q9AII2	
ID Q9AII2	PRELIMINARY; PRT; 381 AA.
AC Q9AII2;	
DT 01-JUN-2001	(TrEMBLrel. 17, Created)
DDT 01-JUN-2001	(TrEMBLrel. 17, Last sequence update)
DE 01-JUN-2001	(TrEMBLrel. 17, Last annotation update)
OS MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).	
GN OMPA.	
OS Chlamydomophila pecorum.	
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomphila.	
OX NCBI_TaxID=85991;	
RN [1]	
RP SEQUENCE OF 34-369 FROM N.A.	
RC STRAIN=1710S;	
RR MEDLINE=93123168; PubMed=8419295;	
XA kaltenboeck B., Kousoulas K.G., Storz J.;	
RT "Structures of and allelic diversity and relationships among the major	
RT outer membrane protein (ompA) genes of the four chlamydial species.";	
RL J. Bacteriol. 175:487-502(1993).	
RN [2]	
RP SEQUENCE FROM N.A.	
RC STRAIN=1710S;	
RR MEDLINE=21078680; PubMed=11211261;	
XA Bush R.M., Everett K.D.;	
RT "Molecular evolution of the Chlamydiaceae.";	
RT Int. J. Syst. Evol. Microbiol. 51:203-220(2001).	
RL [3]	
RP SEQUENCE FROM N.A.	
RC STRAIN=1710S;	
RR Everett K.D.E., Hambly W.A., Andersen A.A.;	
XA Submitted (MAY-2000) to the EMBL/GenBank/DBJ databases.	
RL ENBL; AF269279; AAAK00260.1;	

```

KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 15
FT CHAIN 16 381
SQ SEQUENCE 381 AA; 41332 MW; 29406725CF9D3512 CRC64;

Query Match
Best Local Similarity 70.4%; Score 1480; DB 2; Length 381;
Matches 267; Conservative 47; Mismatches 51; Indels 2; Gaps 2;

QY 2 LPVGNPSPSLLDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDPAKTF 61
DB 16 LPVGNPSPSLLDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDPAKTF 75

QY 62 SWGAKPTG--SAANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVCTLGASN 120
DB 76 SWGAAPTSPDNFADSTVTKQERONPQYKGMHDAEFTNAGYIALNIWDRFVCTLGAS 135

QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYTDTSFWSVGARGALWECG 180
DB 136 GYFKGNSSFNILGILIGISSSTLN-DKLPNANISNGVVELYTDTSFWSVGARGALWECG 194

QY 181 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGKTSAT 240
DB 195 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGKTSAT 254

QY 241 INYHEWVGASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 300
DB 255 INYHEWVGASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSLLG 314

QY 301 NATALSTTDSFDMQIVSQINKFKSRKACGTVGATLVLDADKWSLTAEARLINERAH 360
DB 315 QATQVDTNILDALQIVSLQINKFKSRKACGTVGATLVLDADKWSLTAEARLINERAH 374

QY 361 VSGQRF 367
DB 375 LSAQRF 381

RESULT 9
Q9ALJ5 PRELIMINARY; PRT; 390 AA.
ID Q9ALJ5
AC Q9ALJ5;
DT 01-JUN-2001 (TremBLrel. 17, Created)
DT 01-JUN-2001 (TremBLrel. 17, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NEW JERSEY 1, NJ1;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269266; AAK00247.1; -.
KW Signal.
FT NON_TER 1 1
FT SIGNAL <1 20
FT CHAIN 21 390
SQ SEQUENCE 390 AA; 42042 MW; B62858403DBFAE6 CRC64;

Query Match
Best Local Similarity 70.3%; Score 1479; DB 2; Length 390;
Matches 274; Conservative 41; Mismatches 49; Indels 8; Gaps 4;

QY 2 LPVGNPSPSLLDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDPAKTF 61
DB 16 LPVGNPSPSLLDGTINWGAAGDPCDPCATWCDALISLRAGFYGVDFRILKVDPAKTF 75

QY 62 SWGAKPTGSA--AANYTTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFVCTLGA 118
DB 83 SCMAAAPTAACTASNTVAADRSNFAYGKHLQDAEWCTNAYALNIWDRFVCTLGA 142

QY 119 SNGYTRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYTDTSFWSVGARGALWE 178
DB 143 SNGYTRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVELYTDTSFWSVGARGALWE 201

QY 179 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGCT 236
DB 202 CGCATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAPLPDAGVATATGCT 261

```



```

RESULT 13
Q9X717
ID Q9X717 PRELIMINARY; PRT; 341 AA.
AC Q9X717
DT 01-NOV-1999 (TremBrel. 12, Created)
DT 01-NOV-1999 (TremBrel. 12, Last sequence update)
DT 01-JUN-2001 (TremBrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LW508;
RX MEDLINE=93123168; PubMed=8419295;
RA Kaltenboeck B., Kousoulas K.G., Storz J.;
RT "Structures of and allelic diversity and relationships among the major
RT outer membrane protein (OMP) genes of the four chlamydial species.";
RL J. Bacteriol. 175:487-502(1993).
DR EMBL: M73040; AAD29103.1; -
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam: PF01308; Chlamydia_OMP; 1.
DR ProDom: PD001717; Chlamydia_OMP; 1.
FT NON_TER 1
FT CHAIN 341
FT SIGNAL 1
SQ SEQUENCE 341 AA; 36762 MW; B5933C9BF6AAF171 CRC64;

Query Match 67.5%; Score 1420.5; DB 2; Length 341;
Best Local Similarity 75.4%; Pred. No. 6.4e-107;
Matches 257; Conservative 35; Mismatches 48; Indels 1; Gaps 1;
QY 16 GTIWEAGAGDPCDCAITSLRAGFYGVDFRILKVDAPKTFSGMAKPTGSAAN 74
Db 1 GTIWEAGAGDPCDCAITSLRAGFYGVDFRILKVDAPKTFSGMAKPTGSAAN 74
QY 75 YTTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFVCTLGASNGYIRGNSTAFNLVG 134
Db 1 YTTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFVCTLGASNGYIRGNSTAFNLVG 134
QY 135 LFGVGGTIVNANELNVSNGVLYTDTFSWSVGARGALWECCGATLGAEFYAOASK 194
Db 1 LFGVGGTIVNANELNVSNGVLYTDTFSWSVGARGALWECCGATLGAEFYAOASK 194
QY 121 LIGVKGSSIAADQLPNVGTQIVFYDTTFSWSVGARGALWECCGATLGAEFYAOASN 180
Db 1 LIGVKGSSIAADQLPNVGTQIVFYDTTFSWSVGARGALWECCGATLGAEFYAOASN 180
QY 195 PKVEELNVCNVSQFVNKPKYKGVAPLPDAGVATATGKATINHEWQVGSLSY 254
Db 1 PKVEELNVCNVSQFVNKPKYKGVAPLPDAGVATATGKATINHEWQVGSLSY 254
QY 181 PXIEMLVNVSQAQFVHKPRGYKGTAPLPATGDTQATDTSATIKYHEWQVGLAUSY 240
Db 1 PXIEMLVNVSQAQFVHKPRGYKGTAPLPATGDTQATDTSATIKYHEWQVGLAUSY 240
QY 255 RLSNLVYPYGVWSRATFDADNIRIAQPKLPTAVLNLTWNPSSLGNATLSTTDSFSD 314
Db 255 RLSNLVYPYGVWSRATFDADNIRIAQPKLPTAVLNLTWNPSSLGNATLSTTDSFSD 314
QY 241 RLNLVYPYGVNWSRATFDADNIRIAQPKLPTAVLNLTWNPSSLGNATLSTTDSFSD 300
Db 241 RLNLVYPYGVNWSRATFDADNIRIAQPKLPTAVLNLTWNPSSLGNATLSTTDSFSD 300
QY 315 MQIVSCQINKFKSRKACGVTVGCATLVADKSLTAEARLIN 355
Db 315 MQIVSCQINKFKSRKACGVTVGCATLVADKSLTAEARLIN 355
QY 301 LQIASIQINKMKSRKACGVAVGATLIDADKWSITGEARLIN 341
Db 301 LQIASIQINKMKSRKACGVAVGATLIDADKWSITGEARLIN 341

RESULT 14
Q46193
ID Q46193 PRELIMINARY; PRT; 402 AA.
AC Q46193
DT 01-NOV-1996 (TremBrel. 01, Created)
DT 01-NOV-1996 (TremBrel. 01, Last sequence update)
DT 01-JUN-2001 (TremBrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMP.
OS Chlamydia psittaci (Chlamydia. psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.

RX MEDLINE=94103736; PubMed=8277245;
RA Storey C., Lusher M., Yates P., Richmond S.;
RT "Evidence for Chlamydia pneumoniae of non-human origin.";
RL J. Gen. Microbiol. 139:2621-2626(1993).
DR EMBL: L04980; AAA17396.1; -
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam: PF01308; Chlamydia_OMP; 1.
DR PRINTS: PR01334; CHLAMIDIACOMP.
DR ProDom: PD001717; Chlamydia_OMP; 1.
KW Signal.
FT SIGNAL 1 22 POTENTIAL.
FT CHAIN 23 402 MAJOR OUTER MEMBRANE PROTEIN.
FT SEQUENCE 402 AA; 43261 MW; E36ABC5AF04820A CRC64;

Query Match 67.5%; Score 1419; DB 2; Length 402;
Best Local Similarity 70.5%; Pred. No. 1.1e-106;
Matches 268; Conservative 37; Mismatches 61; Indels 14; Gaps 6;
QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDCAITSLRAGFYGVDFRILKVDAPKTF 61
Db 2 LPVGNPSPSLIDGTIWEAGAGDPCDCAITSLRAGFYGVDFRILKVDAPKTF 61
QY 23 LPVGNPAEPSLLIDGTIWEAGAGDPCDCAITSLRAGFYGVDFRILKVDVNTKF 82
Db 23 LPVGNPAEPSLLIDGTIWEAGAGDPCDCAITSLRAGFYGVDFRILKVDVNTKF 82
QY 62 S-MGAKP--TGSAA-ANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFVCTL 116
Db 62 S-MGAKP--TGSAA-ANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDRFVCTL 116
QY 83 SCMAATPTQATGNASNTNOPEANGRPNIAYGRHMDAEWFSNAFLALNIWDRFVCTL 142
Db 83 SCMAATPTQATGNASNTNOPEANGRPNIAYGRHMDAEWFSNAFLALNIWDRFVCTL 142
QY 117 GASNGYIRGNSTAFNLVGLFGVKGTIVNANE----LPNVSLSNGVVELYTDTFSWSVGA 172
Db 117 GASNGYIRGNSTAFNLVGLFGVKGTIVNANE----LPNVSLSNGVVELYTDTFSWSVGA 172
QY 143 GASNGYIRGNSTAFNLVGLFGVKGTIVNANE----LPNVSLSNGVVELYTDTFSWSVGA 202
Db 143 GASNGYIRGNSTAFNLVGLFGVKGTIVNANE----LPNVSLSNGVVELYTDTFSWSVGA 202
QY 173 RGALWECCGATLGAEFYAOASKPKVEELNVCNVSQFVNKPKYKGVAPLPDAGV 230
Db 173 RGALWECCGATLGAEFYAOASKPKVEELNVCNVSQFVNKPKYKGVAPLPDAGV 230
QY 203 RGALWECCGATLGAEFYAOASKPKVEELNVCNVSQFVNKPKYKGVAPLPDAGV 262
Db 203 RGALWECCGATLGAEFYAOASKPKVEELNVCNVSQFVNKPKYKGVAPLPDAGV 262
QY 231 ATATGKTSATINHEWQVGSLSYRLNLSLVYPYGVWSRATFDADNIRIAQPKLPTAVLN 290
Db 231 ATATGKTSATINHEWQVGSLSYRLNLSLVYPYGVWSRATFDADNIRIAQPKLPTAVLN 290
QY 263 TEATDTKTSATIKYHEWQVGLALSRLNLSLVYPYGVWSRATFDADNIRIAQPKLSEIN 322
Db 263 TEATDTKTSATIKYHEWQVGLALSRLNLSLVYPYGVWSRATFDADNIRIAQPKLSEIN 322
QY 291 LTAWNPSLLGNATLSTT--DSFSDMQIVSCQINKFKSRKACGVTVGCATLVADKWSL 347
Db 291 LTAWNPSLLGNATLSTT--DSFSDMQIVSCQINKFKSRKACGVTVGCATLVADKWSL 347
QY 333 ITTNWPSLLGNATLSTT--DSFSDMQIVSCQINKFKSRKACGVTVGCATLVADKWSL 382
Db 333 ITTNWPSLLGNATLSTT--DSFSDMQIVSCQINKFKSRKACGVTVGCATLVADKWSL 382
QY 348 TAEARLINERAHVSGQFRF 367
Db 348 TAEARLINERAHVSGQFRF 367
QY 383 TGEARLINERAHVNAQFRF 402
Db 383 TGEARLINERAHVNAQFRF 402

RESULT 15
Q9AIJ9
ID Q9AIJ9 PRELIMINARY; PRT; 382 AA.
AC Q9AIJ9
DT 01-JUN-2001 (TremBrel. 17, Created)
DT 01-JUN-2001 (TremBrel. 17, Last sequence update)
DT 01-JUN-2001 (TremBrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia. psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MENINGOPNEUMONITIS, MN, ATCC VR122;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL: AF269262; AAK00243.1; -
KW Signal.
FT NON_TER 1 1 POTENTIAL.
FT SIGNAL <1 2 MAJOR OUTER MEMBRANE PROTEIN.
FT CHAIN 3 382
FT SEQUENCE 382 AA; 41231 MW; 6917171A9A69303B CRC64;
```

Search completed: February 7, 2002, 21:41:21
Job time: 4650 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:34:40 ; Search time 137.02 Seconds
(without alignments)
198.401 Million cell updates/sec

Title: US-09-391-606-16

Perfect score: 1952

Sequence: 1 MLPVGNPDPSPDLDITIME.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A_Geneseq_1101.*

- 1: /SIDS2/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SIDS2/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SIDS2/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SIDS2/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SIDS2/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SIDS2/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SIDS2/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SIDS2/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SIDS2/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SIDS2/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SIDS2/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SIDS2/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SIDS2/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SIDS2/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SIDS2/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SIDS2/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SIDS2/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SIDS2/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SIDS2/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SIDS2/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SIDS2/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SIDS2/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1925	98.6	391	20	AAV56763
2	1647	84.4	343	20	AAV56764
3	1534.5	78.6	389	20	AAV56765
4	1408	72.1	402	20	AAV56766
5	1408	72.1	525	21	AAV56767
6	1408	72.1	525	22	AAV56768
7	1407	72.1	402	20	AAV56769
8	1356.5	69.5	387	20	AAV56770
9	1349	69.1	404	20	AAV56771
10	1345.5	68.9	376	19	AAV56772
11	1338.5	68.6	397	20	AAV56773

12	1336.5	68.5	397	20	AAV56763
13	1326	67.9	396	20	AAV56764
14	1321.5	67.7	397	20	AAV56765
15	1320.5	67.6	376	19	AAV56766
16	1313.5	67.3	393	20	AAV56767
17	1312.5	67.2	393	20	AAV56768
18	1308.5	67.0	372	19	AAV56769
19	1308.5	66.9	393	22	AAV56770
20	1306.5	66.9	374	19	AAV56771
21	1305.5	66.9	395	20	AAV56772
22	1300	66.6	373	19	AAV56773
23	1295	66.3	394	18	AAV56774
24	1295	66.3	394	19	AAV56775
25	1294	66.3	394	21	AAV56776
26	1294	66.3	394	19	AAV56777
27	1283	66.2	394	20	AAV56778
28	1289	66.0	392	20	AAV56779
29	1287	65.9	394	20	AAV56780
30	1284	65.8	394	20	AAV56781
31	1280	65.6	394	7	AAV56782
32	1264.5	64.8	343	20	AAV56783
33	1248	63.9	457	21	AAV56784
34	1214.5	62.2	484	21	AAV56785
35	1182	60.6	356	20	AAV56786
36	1134.5	58.1	514	21	AAV56787
37	1085.5	55.6	454	21	AAV56788
38	1053	53.9	349	21	AAV56789
39	913	46.8	228	20	AAV56790
40	907	46.5	225	20	AAV56791
41	842.5	43.2	222	20	AAV56792
42	825.5	42.3	215	20	AAV56793
43	767	39.3	277	21	AAV56794
44	519	26.6	100	20	AAV56795
45	489	25.1	100	20	AAV56796

ALIGNMENTS

RESULT 1

AAV53319
ID AAV53319 standard; Protein; 391 AA.

AC AAV53319;

DT 13-SEP-1999 (first entry)

DE Chlamydia pneumoniae transmembrane protein sequence.

Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis; sinusitis; purulent otitis media; erythema nodosum; pharyngitis; vaccine; neutralising epitope.

OS Chlamydia pneumoniae.

PN WO9927105-A2.

PD 03-JUN-1999.

PF 20-NOV-1998; 98WO-IB01890.

PR 04-NOV-1998; 98US-0107078.

PR 21-NOV-1997; 97FR-0014673.

(GEST) GENSET.

PI Griffiths R;

WPI; 1999-357842/30.

Genome sequence of Chlamydia pneumoniae

Page 1130-1131; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX SQ Sequence 391 AA;

Query Match 98.6%; Score 1925; DB 20; Length 391;
Best Local Similarity 98.9%; Pred. No. 2e-182; Indels 0; Gaps 0;
Matches 362; Conservative 0; Mismatches 4; Indels 0; Gaps 0;
QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFDRILKVDAPKTF 61
Db 26 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFDRILKVDAPKTF 85
QY 62 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEWTNAGFTALNIWDRFDVFTLGASNG 121
Db 86 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEWTNAGFTALNIWDRFDVFTLGASNG 145
QY 122 YIRGNSTAFNLVGLFGVKTNNANLNPVSLNSGVVELYDTTSFWSVGARGALWECGC 181
Db 146 YIRGNSTAFNLVGLFGVKTNNANLNPVSLNSGVVELYDTTSFWSVGARGALWECGC 205
QY 182 ATLGAEFOYQAQSKPKVEELNVCNVSQFVNPKGKYGVAFLPTDAGVATATGKTSATI 241
Db 206 ATLGAEFOYQAQSKPKVEELNVCNVSQFVNPKGKYGVAFLPTDAGVATATGKTSATI 265
QY 242 NYHEWQVGASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 301
Db 266 NYHEWQVGASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 325
QY 302 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 361
Db 326 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 385
QY 362 SQGFRF 367
Db 386 SQGFRF 391

RESULT 2
AAY56771
ID AAY56771 standard; Protein; 343 AA.

XX AC AAY56771;
XX DT 22-FEB-2000 (first entry)
XX DE C. trachomatis serovar HuPn MOMP sequence.

XX KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
XX KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
XX OS Chlamydia trachomatis.

XX PN W09951745-A2.
XX PD 14-OCT-1999.
XX PF 07-APR-1999; 99WO-CA00292.
XX PR 07-APR-1998; 98US-0055765.
XX PA (UYMA-) UNIV MANITOBA.

XX Bruham RC;
XX WPI; 1999-620205/53.
XX Non-replicating vector encoding fragments of the outer membrane protein
XX of Chlamydia, useful in vaccines and as immunogen -
XX Disclosure; Fig 10 A-F; 52pp; English.
CC The invention provides a non-replicating vector that comprises, linked
CC to a promoter, a nucleotide sequence that encodes a region containing at
CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
CC vaccines to generate a protective immune response (mainly cellular)
CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
CC in standard immunoassays. Immunization with the vector induces a broad
CC spectrum of immune responses, including Th1-like CD4 responses and
CC mucosal immunity, providing significant protection against subsequent
CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
CC of serovars of C. trachomatis.
XX SQ Sequence 343 AA;

Query Match 84.4%; Score 1647; DB 20; Length 343;
Best Local Similarity 86.3%; Pred. No. 6.4e-155;
Matches 316; Conservative 2; Mismatches 2; Indels 46; Gaps 1;
QY 2 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFDRILKVDAPKTF 61
Db 24 LPVGNPSPSLIDGTIWEAGAGDPCDPCATWCDATSLRAGFYGDYVDFDRILKVDAPKTF 83
QY 62 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEWTNAGFTALNIWDRFDVFTLGASNG 121
Db 84 SMGAKPTGSAANYTAVDRPNPYNKHLHDAEWTNAGFTALNIWDRFDVFTLGASNG 143
QY 122 YIRGNSTAFNLVGLFGVKTNNANLNPVSLNSGVVELYDTTSFWSVGARGALWECGC 181
Db 144 YIRGNSTAFNLVGLFGVKTNNANLNPVSLNSGVVELYDTTSFWSVGARGALWECGC 193
QY 182 ATLGAEFOYQAQSKPKVEELNVCNVSQFVNPKGKYGVAFLPTDAGVATATGKTSATI 241
Db 194 ATLGAEFOYQAQSKPKVEELNVCNVSQFVNPKGKYGVAFLPTDAGVATATGKTSATI 217
QY 242 NYHEWQVGASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 301
Db 218 NYHEWQVGASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVNLTAWNPSSLGN 277
QY 302 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 361
Db 278 ATALSTDSFDMQIVSCQINKFKSRKACGVTGATLVADADKWSLTAEARLINERAAHV 337
QY 362 SQGFRF 367
Db 338 SQGFRF 343

RESULT 3

AAW98188
ID AAW98188 standard; Protein; 389 AA.

XX AC AAW98188;
XX DT 05-JUL-1999 (first entry)

XX DE Chlamydia psittaci major outer membrane protein.
XX KW Major outer membrane protein; MOMP; psittacosis; infection;
XX KW vaccine; genetic immunisation.
XX OS Chlamydia psittaci.
XX

PN W09910005-A1.
 XX
 PD 04-MAR-1999.
 XX
 PF 28-AUG-1998; 98WO-US17943.
 XX
 PR 28-AUG-1997; 97US-0057147.
 XX
 PA (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
 XX
 PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
 XX
 DR WPI; 1999-254214/21.
 XX
 XX N-PSDB; AAX25047.
 XX
 PT A new vaccine for Chlamydia psittaci infections
 XX
 PS Disclosure; Page 60-61; 72pp; English.
 XX
 CC The present sequence is the major outer membrane protein (MOMP)
 CC of Chlamydia psittaci strain B577. A claimed MOMP polypeptide (see
 CC AAW98184) comprises regions VD3 and VD4 of B577 MOMP, i.e. it lacks
 CC regions VD1 and VD2. A claimed vaccine composition includes MOMP
 CC polypeptide lacking VD1 and VD2, optionally fused to a maltose
 CC binding protein. Also claimed are an isolated nucleic acid
 CC encoding the polypeptide, a vector, and a method of preventing C.
 CC psittaci infection by administering the vaccine containing the
 CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking
 CC regions VD1 and VD2 are useful for genetic vaccination. The
 CC vaccines are used to prevent C. psittaci infection, especially in
 CC birds.
 XX
 SQ Sequence 389 AA;
 Query Match 78.68; Score 1534.5; DB 20; Length 389;
 Best Local Similarity 75.78; Pred. No. 1.1e-143;
 Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;
 QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
 DB 23 lfvgnpaeapsllidgtmwegasgdpdpcatwcdaaisragsygydvfrlvkdvntki 82
 QY 62 S-MGAKPTGSAANTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTPLGASN 120
 DB 83 tqmgavptgtaaanvktptdrpnlayghlqdaeftnaafalnwiwdrfdifctlgasn 142
 QY 121 GYIRGNSTAFNLVGLFGVKGTTVNANLNPVSLNSGVVELYTDTSFSVSGARGALWECG 180
 DB 143 gyfkassaafnlvlgvkgssiaadqipnvgitggyefytdtcfsvsgargalwecg 202
 QY 181 CATLGAEFQYAOQSKPVEELNVCNVSQFSVKNPKGYGVAFPLPTDAGVATATGTSAT 240
 DB 203 catlgaefqyagsnkiemlnvsspaqfvvhkprgykgtafplptagtdqatdtksat 262
 QY 241 INYHEMVOYASLSYRLNSLVPYIGVOWSRATFDADNIRIAQPKLPTAVLNTAWNPSLLG 300
 DB 263 lkyhewqvgalsyrlnmlvpyisvwnsratfdadairiaqpklaaavlnittwnptllg 322
 QY 301 NATALSTTDSFDMQIVSCQINPKFKRACGVTVGATLVADKWSLTAEARLINERAH 360
 DB 323 eataldtsnkfdafqiasiqinkmksrkacgvavgatlidadkwsitgearlineraah 382
 QY 361 VSGQFRF 367
 DB 383 mnaqfrf 389
 RESULT 4
 AAW98189
 ID AAW98189 standard; Protein: 402 AA.
 XX
 AC AAW98189;

XX 05-JUL-1999 (first entry)
 DT
 XX Chlamydia psittaci major outer membrane protein.
 DE
 XX Major outer membrane protein; MOMP; psittacosis; infection;
 KW vaccine; genetic immunisation.
 XX
 OS Chlamydia psittaci.
 XX
 PN W09910005-A1.
 XX
 PD 04-MAR-1999.
 XX
 PF 28-AUG-1998; 98WO-US17943.
 XX
 PR 28-AUG-1997; 97US-0057147.
 XX
 PA (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
 XX
 PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
 XX
 DR WPI; 1999-254214/21.
 XX
 XX N-PSDB; AAX25048.
 XX
 PT A new vaccine for Chlamydia psittaci infections
 XX
 PS Disclosure; Page 65-66; 72pp; English.
 XX
 CC The present sequence is the major outer membrane protein (MOMP)
 CC of Chlamydia psittaci strain 6BC. Claimed MOMP polypeptides (see
 CC AAW98183 and AAW98184) comprise regions VD3 and VD4 of an MOMP, i.e.
 CC they lack regions VD1 and VD2. Claimed vaccine compositions
 CC include such MOMP polypeptides, optionally fused to a maltose
 CC binding protein. Also claimed are isolated nucleic acids encoding
 CC the polypeptide, a vector, and a method of preventing C. psittaci
 CC infection by administering the vaccine containing the MOMP
 CC polypeptide. Vectors encoding MOMP polypeptides lacking regions
 CC VD1 and VD2 are useful for genetic vaccination. The vaccines are
 CC used to prevent C. psittaci infection, especially in birds.
 XX
 SQ Sequence 402 AA;
 Query Match 72.18; Score 1408; DB 20; Length 402;
 Best Local Similarity 70.38; Pred. No. 4.2e-131;
 Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;
 QY 2 LPVGNPSPSLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
 DB 23 lfvgnpaeapsllidgtmwegasgdpdpcatwcdaaisragsygydvfrlvkdvntkf 82
 QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVDFCTL 116
 DB 83 sgmaatpqtatgnasntnqpeangrpnlaygrhmqdaefnaafalnwiwdrfdictl 142
 QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVNANLNPVSLNSGVVELYTDTSFSWSVGA 172
 DB 143 gasngyfkassaafnlvlgfisaassistdlpmqlpnvgitggyvfevtdtsfswsya 202
 QY 173 RGALEWCCCATLGAEFQYAOQSKPVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGV 230
 DB 203 rgalewcccatlgaefqyagsnkiemlnvsspaqfvvhkprgykgtassnfppltagt 262
 QY 231 ATATGTSATINYHEMVOYASLSYRLNSLVPYIGVOWSRATFDADNIRIAQPKLPTAVLN 290
 DB 263 teatdtksatikyhewqvgalsyrlnmlvpyisvwnsratfdadairiaqpklskein 322
 QY 291 LTANNPSLLGNATAL---STTDSFDMQIVSCQINPKFKRACGVTVGATLVADKWSL 347
 DB 323 lttwnpsllgstaltalpnnsqkdvlsdvliqiasiqinkmksrkacgvavgatlidadkws 382
 QY 348 TAEARLINERAHVSGQFRF 367

Query Match 72.1%; Score 1408; DB 22; Length 525;
Best Local Similarity 70.3%; Pred. No. 6.3e-131;
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDAPKTF 61
DB 146 LPVGNPAEPSLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDNKTf 205

QY 62 S-MGAKPT---GSAA-ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116
DB 206 sgmaatptqatgnasntngpeangrpnlaygrhmqdaefwnaafinaidwrfdiycl 265

QY 117 GASNGYIRGNSTAFNLVGLFGVKG----TTVNANELPNVSLNGVVELYDTDFSNVSGA 172
DB 266 gasngyfkassafnlvlgfsgaassistdipmqlpnvgitqgvvfytdtsfswsvga 325

QY 173 RGALWECGCATIGAEOYQAQSKPKVEELNVICNVSOFSVKNPKYKGVNA--PFLPTDAGV 230
DB 326 rgalwecgcataigaefqyqasnpkilemInvtspsqafvlnkprgykgaasnfpilptagt 385

QY 231 ATATGKTSATINYHEWQVGSLSYRLNSLPVYIGVQWSRATEDADNIRIAQPKLPTAVLN 290
DB 386 teatdtksatikyheqwglalsyrlnmlypyigvnwstratfdadtirraqpkkseilin 445

QY 291 LTAWNPSLLGNATAL---STTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADAKWSL 347
DB 446 itwnpslligsttalpnnsqkdvlsdvqlasiginkmksrkacgvavgatliadadkwsli 505

QY 348 TAEARLINERAHVSGQFRF 367
DB 506 tgearlineraahmaaqfrf 525

RESULT 7

ID AAW98187
XX AAW98187 standard; Protein; 402 AA.
AC AAW98187;
DT 05-JUL-1999 (first entry)
XX Chlamydia psittaci major outer membrane protein.
DE Major outer membrane protein; MOMP; psittacosis; infection;
KW vaccine; genetic immunisation.
XX Chlamydia psittaci.
OS WO9910005-A1.
PN 04-MAR-1999.
XX 28-AUG-1998; 98WO-US17943.
XX 28-AUG-1997; 97US-0057147.
XX (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
XX WPI; 1999-254214/21.
DR N-PSDB; AAX25046.
XX A new vaccine for Chlamydia psittaci infections
PT Disclosure; Page 56-57; 72pp; English.
PS The present sequence is the major outer membrane protein (MOMP)
CC of Chlamydia psittaci strain LSUWTK, a cocktail isolate (the
CC MOMP gene sequence of this isolate is identical to that of C.
CC psittaci Avian Type C). A claimed MOMP polypeptide (see AAW98183)
CC comprises regions VD3 and VD4 of LSUWTK MOMP, i.e. it lacks
CC regions VD1 and VD2. A claimed vaccine composition includes MOMP

CC polypeptide lacking VD1 and VD2, optionally fused to a maltose
CC binding protein. Also claimed are an isolated nucleic acid
CC encoding the polypeptide, a vector, and a method of preventing C.
CC psittaci infection by administering the vaccine containing the
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking
CC regions VD1 and VD2 are useful for genetic vaccination. The
CC vaccines are used to prevent C. psittaci infection, especially in
CC birds.

XX Sequence 402 AA;
SQ

Query Match 72.1%; Score 1407; DB 20; Length 402;
Best Local Similarity 70.0%; Pred. No. 5.3e-131;
Matches 266; Conservative 39; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSLLIDGTWEGAGDPCDPCATWCDALSLRAGFYGDYVDFRILKVDNKTf 82

QY 62 S-MGAKP---TGSA-ANYTTAVDRPNPAYNKHLDHAEFTNAGFTALNIWDRFVFCFL 116
DB 83 sgmaatptqatgnasntngpeangrpnlaygrhmedaefwnaafinaidwrfdiycl 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG----TTVNANELPNVSLNGVVELYDTDFSNVSGA 172
DB 143 gasngyfkassafnlvlgfsgaassistdipmqlpnvgitqgvvfytdtsfswsvga 202

QY 173 RGALWECGCATIGAEOYQAQSKPKVEELNVICNVSOFSVKNPKYKGVNA--PFLPTDAGV 230
DB 203 rgalwecgcataigaefqyqasnpkilemInvtspsqafvlnkprgykgaasnfpilptagt 262

QY 231 ATATGKTSATINYHEWQVGSLSYRLNSLPVYIGVQWSRATEDADNIRIAQPKLPTAVLN 290
DB 263 teatdtksatikyheqwglalsyrlnmlypyigvnwstratfdadtirraqpkkseilin 322

QY 291 LTAWNPSLLGNATAL---STTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADAKWSL 347
DB 323 itwnpslligsttalpnnsqkdvlsdvqlasiginkmksrkacgvavgatliadadkwsli 382

QY 348 TAEARLINERAHVSGQFRF 367
DB 383 tgearlineraahmaaqfrf 402

RESULT 8

AAV56767
ID AAV56767 standard; Protein; 387 AA.
XX AAV56767;
AC AAV56767;
DT 22-FEB-2000 (first entry)
XX C. trachomatis serovar MoPh MOMP sequence.
DE Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
XX Chlamydia trachomatis.
OS WO9951745-A2.
PN 14-OCT-1999.
XX 07-APR-1999; 99WO-CA00292.
XX 07-APR-1998; 98US-0055765.
XX (UYMA-) UNIV MANITOBA.
PI Bruhnam RC;
XX WPI; 1999-620205/53.

KW Major outer membrane protein; antibody; antigen; diagnosis; disease.
 XX Chlamydia trachomatis.
 OS JP10234395-A.
 PN 08-SEP-1998.
 PD 25-FEB-1997; 97JP-0040780.
 PF 25-FEB-1997; 97JP-0040780.
 PR (ELED) DENKI KAGAKU KOGYO KK.
 XX WPI; 1998-535045/46.
 DR Preparation of major outer membrane protein of Chlamydia trachomatis
 PT - by cloning and recombinant expression of the gene, for use as a
 PT diagnostic of Chlamydia infections
 XX
 PS Claim 2; Page 15-16; 19pp; Japanese.
 XX AAW76362-W76366 are major outer membrane proteins isolated from
 CC Chlamydia trachomatis which are used for the measurement of an antibody
 CC against Chlamydia trachomatis, by using it as an antigen in the form of
 CC a reagent. The method can give a diagnostic method for Chlamydia
 CC trachomatis-infected diseases with high specificity.
 XX
 SQ Sequence 376 AA;

Query Match 68.9%; Score 1345.5; DB 19; Length 376;
 Best Local Similarity 65.2%; Pred. No. 6.1e-125;
 Matches 245; Conservative 51; Mismatches 71; Indels 9; Gaps 3;
 QY 1 MLPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKT 60
 DB 1 mlpvgnpaepsimldgilwegfggpcdpcatwcdaismrvgygdfvdrvktvdkne 60
 QY 61 FSMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEFTNAGFTALNIDWDFVCTL 116
 DB 61 fmggaaptndaaadlqndpknvarpnaygkqhmgaemftnaaymalniwdrfvctl 120
 QY 117 GASNGYIRGNSTAFNLVGLFGVKGTTVANE---LPNVSLNGVVELYDTDFTSFWSVGAR 173
 DB 121 gattgylkgnasfnlglfgtkssdfntaklvpnlalnrvvlydttdtfawsvgar 180
 QY 174 GALWECGCATLGAEFQYAOQSKPKVEELNVICNVQSFVNKPKYKGVAFPLPTDAGVATA 233
 DB 181 aalwecgcattlgasfgyaqskpkveelnvlnaseftinkpkgyvgaefpldtagtea 240
 QY 234 TGTKSATINYHEWQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVINLTA 293
 DB 241 tgtkdasidyhewqaslsalysrlnmftpylgvkwsvsfadtiriacpklaeaildvtt 300
 QY 294 WNPSSLGNAT--ALSTTDSFDMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAEA 351
 DB 301 lnptiagktvvasgsdndladtmqivslqlnkmksrksclavgttvdadkyavtvvet 360
 QY 352 RLINERAAHVSGQFRF 367
 DB 361 rlideraahvnaqfrf 376

RESULT 11
 ID AAY56766
 XX AAY56766 standard; Protein; 397 AA.
 AC AAY56766;
 XX 22-FEB-2000 (first entry)
 DT
 XX
 DE C. trachomatis serovar H MOMP sequence.

XX Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
 XX Chlamydia trachomatis.
 OS WO9951745-A2.
 PN 14-OCT-1999.
 PD 07-APR-1999; 99WO-CA00292.
 PF 07-APR-1998; 98US-0055765.
 PR (UYMA-) UNIV MANITOBA.
 XX Bruhnam RC;
 PI WPI; 1999-620205/53.
 DR Non-replicating vector encoding fragments of the outer membrane protein
 PT of Chlamydia, useful in vaccines and as immunogen
 XX Disclosure; Fig 10 A-F; 52pp; English.
 PS The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX
 SQ Sequence 397 AA;

Query Match 68.6%; Score 1338.5; DB 20; Length 397;
 Best Local Similarity 64.8%; Pred. No. 3.3e-124;
 Matches 243; Conservative 52; Mismatches 71; Indels 9; Gaps 3;
 QY 2 LPVGNPSDPSLLIDGTWEGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKT 61
 DB 23 lpgvnpaepsimldgilwegfggpcdpcatwcdaismrvgygdfvdrvktvdknef 82
 QY 62 SMGAKPTGSAANY-----TTAVDRPNPAYNKHLDHAEFTNAGFTALNIDWDFVCTL 117
 DB 83 qmgaaptndaaadlqndpknvarpnaygkqhmgaemftnaaymalniwdrfvctlg 142
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTVANE---LPNVSLNGVVELYDTDFTSFWSVGAR 174
 DB 143 atcygylkgnasfnlglfgtkssdfntaklvpnlalnrvvlydttdtfawsvgara 202
 QY 175 ALWECGCATLGAEFQYAOQSKPKVEELNVICNVQSFVNKPKYKGVAFPLPTDAGVATA 234
 DB 203 alwecgcattlgasfgyaqskpkveelnvlnaseftinkpkgyvgaefpldtagtea 262
 QY 235 TGTKSATINYHEWQVQASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVINLTA 294
 DB 263 gtkdasidyhewqaslsalysrlnmftpylgvkwsvsfadtiriacpklaeaildvttl 322
 QY 295 NPSLLGNAT--ALSTTDSFDMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAEA 352
 DB 323 nptiagktvvasgsdndladtmqivslqlnkmksrksclavgttvdadkyavtvvet 382
 QY 353 LINERAAHVSGQFRF 367
 DB 383 rlideraahvnaqfrf 397

RESULT 12
 AAY56763
 ID AAY56763 standard; Protein; 397 AA.
 XX AC AAY56763;
 XX DT 22-FEB-2000 (first entry)
 XX DE C. trachomatis serovar L3 MOMP sequence.
 XX KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 XX KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
 XX OS Chlamydia trachomatis.
 XX PN WO9951745-A2.
 XX PD 14-OCT-1999.
 XX PF 07-APR-1999; 99WO-CA00292.
 XX PR 07-APR-1998; 98US-0055765.
 XX PA (UYMA-) UNIV MANITOBA.
 XX PI Bruham RC;
 XX DR WPI; 1999-620205/53.
 XX PT Non-replicating vector encoding fragments of the outer membrane protein
 XX PT of Chlamydia, useful in vaccines and as immunogen
 XX PS Disclosure; Fig 10 A-F; 52pp; English.
 XX CC The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX SQ Sequence 397 AA;
 Query Match 68.5%; Score 1336.5; DB 20; Length 397;
 Best Local Similarity 64.8%; Pred. No. 5.1e-124; Indels 9; Gaps 3;
 Matches 243; Conservative 50; Mismatches 73;
 QY 2 LPVGNPSPSLIDGTIWEGAAGDCPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
 DB 23 LPVGNPAEPSIMIDGILWEGFGGDCPCCTTWCDAISMRYGGYGFVDFRILKVDNKEF 82
 QY 62 SMGAKPTGSAANY-----TTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDVCTLG 117
 DB 83 QMGAAPTTSDTAGLNDPTNVARPNPAGYGHMGAEMFTNAAYMALNIWDRFDVCTLG 142
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTVNAE---LPNVLSNGVVELYTDTSFWSVGARG 174
 DB 143 ATTGYLKGNSASFLVGLFGVKGTTQSTNFNTAKLPNTALNQAVVELYTDTSFWSVGARG 202
 QY 175 ALWECGCATLGAFFOYQAKPVEELNVCNYSQFVSNKPKYKGYGAFPLPTDAGVATAT 234
 DB 203 ALWECGCATLGAFFOYQAKPVEELNVCNYSQFVSNKPKYKGYGAFPLPTDAGVATAT 262
 QY 235 GTRKATINHEWQVGLASLYRLNSLVPYIGVQVRSRATEDADNIRIAQPKLPTAVLNLTAW 294
 DB 263 GTKDASIDYHEWQVGLASLYRLNSLVPYIGVQVRSRATEDADNIRIAQPKLPTAVLNLTAW 322

QY 295 NPSLL--GNATALSTTDSFDEMOIVSCQINKPKSRKACGVTGATLVLDADKWSLTAEAR 352
 DB 323 nptiagkgsvvsgseneladtmtqlvslqlnkmksrkscgiavgttividadkyavtvetr 382
 QY 353 LINERAAHVSGQERF 367
 DB 383 llderaahvnaqfrf 397
 RESULT 13
 AAY56764
 ID AAY56764 standard; Protein; 396 AA.
 XX AC AAY56764;
 XX DT 22-FEB-2000 (first entry)
 XX DE C. trachomatis serovar A MOMP sequence.
 XX KW Major outer membrane protein; MOMP; Chlamydia; vaccine; immune response;
 XX KW cellular response; immunogen; Th1-like CD4 response; mucosal immunity.
 XX OS Chlamydia trachomatis.
 XX PN WO9951745-A2.
 XX PD 14-OCT-1999.
 XX PF 07-APR-1999; 99WO-CA00292.
 XX PR 07-APR-1998; 98US-0055765.
 XX PA (UYMA-) UNIV MANITOBA.
 XX PI Bruham RC;
 XX DR WPI; 1999-620205/53.
 XX PT Non-replicating vector encoding fragments of the outer membrane protein
 XX PT of Chlamydia, useful in vaccines and as immunogen
 XX PS Disclosure; Fig 10 A-F; 52pp; English.
 XX CC The invention provides a non-replicating vector that comprises, linked
 CC to a promoter, a nucleotide sequence that encodes a region containing at
 CC least one of the conserved domains 2, 3 and 5 of a major outer membrane
 CC protein (MOMP) of a Chlamydia strain. The vector is used: (a) in
 CC vaccines to generate a protective immune response (mainly cellular)
 CC against MOMP, and (b) as immunogens to raise anti-MOMP antibodies, useful
 CC in standard immunoassays. Immunization with the vector induces a broad
 CC spectrum of immune responses, including Th1-like CD4 responses and
 CC mucosal immunity, providing significant protection against subsequent
 CC challenge. Sequences AAY56757-71 represent MOMP sequences from a variety
 CC of serovars of C. trachomatis.
 XX SQ Sequence 396 AA;
 Query Match 67.9%; Score 1326; DB 20; Length 396;
 Best Local Similarity 64.7%; Pred. No. 5.6e-123;
 Matches 242; Conservative 48; Mismatches 76; Indels 3; Gaps 3;
 QY 2 LPVGNPSPSLIDGTIWEGAAGDCPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
 DB 23 LPVGNPAEPSIMIDGILWEGFGGDCPCCTTWCDAISMRYGGYGFVDFRILKVDNKEF 82
 QY 62 SMGAKPTGSAANY-----TTAVDRNPAYNKHLDHAEWFTNAGFTALNIWDRFDVCTLG 117
 DB 83 QMGAAPTTSDTAGLNDPTNVARPNPAGYGHMGAEMFTNAAYMALNIWDRFDVCTLG 142
 QY 118 ASNGYIRGNSTAFNLVGLFGVKGTTV---NANFELPNVLSNGVVELYTDTSFWSVGARG 174
 DB 143 ATTGYLKGNSASFLVGLFGVKGTTQSTNFNTAKLPNTALNQAVVELYTDTSFWSVGARG 202


```
Db 1 mlpvgnpaepslmidgilwegfgdpcpcttwcdaismrvggygdfvdrvlktdvne 60
Qy 61 FPMGAKPTGSAANY---TTAVDRPNPAYNKHLDHAEWFTNAGFIALNIWDREDFECTL 116
Db 61 fmqdaapttsdvaqlqndptinvarpnpaygkhmqdaemftnaaymainiwdrfdvictl 120
Qy 117 GASNGYIRGNSTAFNLVCLFGVKGTVNANE--LPNVSLSNGVVELYTDTSFSWSVGAR 173
Db 121 gattgyikgnssasfnlvglfytktqsssfntaklipntalneavvelyinttfawsvgar 180
Qy 174 GALWECGCATLGAEFOYAQSKPKVEELNVICNVSQFSVKNPKGYKGVAFPLPTDAGVATA 233
Db 181 aalwecgcattlgasfqyaqskpkveelnvlnaseftinkpkgyvgaefpinitagteaa 240
Qy 234 TGTKSATINYHEWQVGASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLT 293
Db 241 tgtkdasidyhewqasialsyrlnmftpyigkwsrvsfdattiriaqpklaeaildvtt 300
Qy 294 WNPSSLGNATALS--TTDSFDEMQIVSCQINKFKSRKACGVTVGATLVADAKWSLTAE 351
Db 301 lnrttagkgsvvsgtdneladtmgivslqinkmksrkscgiavgttivdadkyavtvea 360
Qy 352 RLINERAAHVSGQPRF 367
Db 361 rlideraahvnaqfrf 376
```

Search completed: February 7, 2002, 21:34:40
Job time: 20759 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run On: February 7, 2002, 21:36:25 ; Search time 85.91 seconds
(without alignments)
96.132 Million cell updates/sec

Title: US-09-391-606-16
Perfect score: 1952
Sequence: 1 MLPVGNPDSPLLDGTWE.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A_COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B_COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A_COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B_COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PTUS_COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	136.5	7.0	61	US-08-374-560-1	Sequence 1, Appli
2	125	6.4	42	PCT-US93-08739-3	Sequence 3, Appli
3	124	6.4	25	US-08-446-692-45	Sequence 45, Appl
4	124	6.4	25	US-08-488-351A-45	Sequence 45, Appl
5	124	6.4	25	PCT-US93-08739-1	Sequence 1, Appli
6	124	6.4	25	PCT-US95-13841-17	Sequence 17, Appl
7	124	6.4	37	US-08-446-692-24	Sequence 24, Appl
8	124	6.4	37	US-08-488-351A-24	Sequence 24, Appl
9	119	5.1	31	US-08-513-385B-1	Sequence 1, Appli
10	115	5.9	25	US-09-100-409A-51	Sequence 51, Appl
11	101.5	5.2	919	US-08-788-674-4	Sequence 4, Appli
12	97.5	5.0	457	US-09-142-759-1	Sequence 1, Appli
13	90.5	4.6	784	US-09-371-913A-7	Sequence 7, Appli
14	87.5	4.5	834	US-08-471-033-21	Sequence 21, Appl
15	87.5	4.5	834	US-08-471-044-21	Sequence 21, Appl
16	87.5	4.5	834	US-08-463-483A-21	Sequence 21, Appl
17	87.5	4.5	834	US-08-471-046A-21	Sequence 21, Appl
18	87.5	4.5	834	US-08-470-566B-21	Sequence 21, Appl
19	87.5	4.5	834	US-08-469-334-21	Sequence 21, Appl
20	87.5	4.5	834	US-09-300-529-21	Sequence 21, Appl
21	85	4.4	621	US-08-262-220-8	Sequence 8, Appli
22	85	4.4	621	US-08-471-733-8	Sequence 8, Appli
23	85	4.4	621	US-08-468-878-8	Sequence 8, Appli
24	85	4.4	621	US-08-750-494-8	Sequence 8, Appli
25	85	4.4	628	US-07-952-853-22	Sequence 22, Appl
26	85	4.4	628	US-08-914-848-22	Sequence 22, Appl
27	85	4.4	664	US-09-295-186-17	Sequence 17, Appl

28	81	4.1	306	1	US-08-312-387B-4	Sequence 4, Appli
29	81	4.1	306	1	US-08-683-428-4	Sequence 4, Appli
30	81	4.1	306	1	US-08-683-458-4	Sequence 4, Appli
31	81	4.1	306	2	US-08-878-360-4	Sequence 4, Appli
32	81	4.1	306	3	US-08-478-140B-4	Sequence 4, Appli
33	81	4.1	400	1	US-08-118-906-14	Sequence 14, Appl
34	81	4.1	400	1	US-08-486-196-14	Sequence 14, Appl
35	81	4.1	400	1	US-08-488-135-14	Sequence 14, Appl
36	81	4.1	400	2	US-08-474-063-14	Sequence 14, Appl
37	81	4.1	400	4	US-09-233-506-4	Sequence 4, Appli
38	81	4.1	592	1	US-08-217-327-8	Sequence 8, Appli
39	81	4.1	1338	2	US-08-728-470-9	Sequence 9, Appli
40	81	4.1	1338	4	US-08-719-641-9	Sequence 9, Appli
41	81	4.1	1529	2	US-08-728-470-10	Sequence 10, Appl
42	81	4.1	1529	4	US-08-719-641-10	Sequence 10, Appl
43	81	4.1	1599	2	US-08-617-697-9	Sequence 9, Appli
44	81	4.1	1600	2	US-08-617-697-10	Sequence 10, Appl
45	80.5	4.1	667	1	US-08-471-033-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1
US-08-374-560-1
Sequence 1, Application US/08374560
Patent No. 5882645
GENERAL INFORMATION:
APPLICANT: TOTH, Istvan
APPLICATOR: GIBBONS, William Anthony
TITLE OF INVENTION: PEPTIDE COMPOUNDS
NUMBER OF SEQUENCES: 39
CORRESPONDENCE ADDRESS:
ADDRESSEE: Ostrolenk, Faber, Gerb & Soffen
STREET: 1180 Avenue of the Americas
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10036-8403
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/374,560
FILING DATE: 13-MAR-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9215780.9
FILING DATE: 24-JUL-1992
ATTORNEY/AGENT INFORMATION:
NAME: Meilman, Edward A.
REGISTRATION NUMBER: 24,735
REFERENCE/DOCKET NUMBER: P/365-302
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 382-0700
TELEFAX: (212) 382-0888
TELEX: 236925
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 61 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: peptide
FRAGMENT TYPE: internal
US-08-374-560-1

Query Match 7.0%; Score 136.5; DB 2; Length 61;
Best Local Similarity 47.5%; Pred. No. 3.2e-07;
Matches 29; Conservative 10; Mismatches 21; Indels 1; Gaps 1;

```
QY 269 RATEADNIRIAQKPLPAVLNLTAWNPISLGNATALSTTD-SFSDFMQIVSCQINKFKS 327
      ||||||| ||||||| : :|| |||: | : : : | ||||| :||| ||
Db 1 RASFDADTIRIAQPKSAETIFDVTTLPTIAGAGDVKTSABGQLGDTMQIVSLQLNKMK 60
      ||||||| ||||||| : :|| |||: | : : : | ||||| :||| ||
QY 328 R 328
      |
Db 61 R 61

RESULT 2
PCT-US93-08739-3
; Sequence 3, Application PC/TUS9308739
; GENERAL INFORMATION:
; APPLICANT: The Government of the United States of America
; APPLICANT: as represented
; APPLICANT: by the Secretary of the Department of Health and Human Services
; TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR
; TITLE OF INVENTION: CHLAMYDIA
; TITLE OF INVENTION: TRACHOMATIS
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US93/08739
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/947,671 US
; FILING DATE: 18 SEP 93
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 42 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
PCT-US93-08739-3

Query Match 6.4%; Score 125; DB 5; Length 42;
Best Local Similarity 61.9%; Pred. No. 2.8e-06;
Matches 26; Conservative 4; Mismatches 2; Indels 10; Gaps 2;

QY 103 ALNIWDRFDVFTLGASNGYIRGNSTAFNLVLGLFGVKGTTVN 144
      ||||||| ||||||| : :|| |||: | : : : | ||||| :||| ||
Db 1 ALNIWDRFDVFTLGATGTLKAGNS-----FDV--TTLN 32

RESULT 3
US-08-446-692-45
; Sequence 45, Application US/08446692
; Patent No. 5759551
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053

Query Match 6.4%; Score 124; DB 1; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
      ||||||| ||||||| : :|| |||: | : : : | ||||| :||| ||
Db 1 ALNIWDRFDVFTLGATGTLKAGNS 25

RESULT 4
US-08-488-351A-45
; Sequence 45, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,351A
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/446,692
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,275
FILING DATE: 14-APR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,166
FILING DATE: 27-APR-1992
CLASSIFICATION: 424
```

```
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/446,692
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Maria C.H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146 US2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212)415-8745
TELEFAX: (516)751-6849
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-446-692-45

Query Match 6.4%; Score 124; DB 1; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
      ||||||| ||||||| : :|| |||: | : : : | ||||| :||| ||
Db 1 ALNIWDRFDVFTLGATGTLKAGNS 25
```

```
RESULT 4
US-08-488-351A-45
; Sequence 45, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: and synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/488,351A
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/446,692
FILING DATE: 7-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/229,275
FILING DATE: 14-APR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/057,166
FILING DATE: 27-APR-1992
CLASSIFICATION: 424
```

ATTORNEY/AGENT INFORMATION:
NAME: Maria C.H. Lin
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4146 US2
TELEPHONE: (212)415-8745
TELEFAX: (516)751-6849
INFORMATION FOR SEQ ID NO: 45:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-488-351A-45

Query Match 6.4%; Score 124; DB 2; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127
|||||
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 5

PCT-US93-08739-1
Sequence 1, Application PC/TUS9308739
GENERAL INFORMATION:
APPLICANT: The Government of the United States of America
APPLICANT: as represented
TITLE OF INVENTION: SYNTHETIC PEPTIDE VACCINE FOR
TITLE OF INVENTION: CHLAMYDIA
TITLE OF INVENTION: TRACHOMATIS
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/08739
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/947,671 US
FILING DATE: 18 SEP 93
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
PCT-US93-08739-1

Query Match 6.4%; Score 124; DB 5; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127
|||||
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 6

PCT-US95-13841-17
Sequence 17, Application PC/TUS9513841
GENERAL INFORMATION:

APPLICANT: United Biomedical Inc; Walfield, Alan M.;
APPLICANT: Wang, Chang Yi
TITLE OF INVENTION: Synthetic IgE Membrane Anchor
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10154
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: WordPerfect 5.1
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/13841
FILING DATE: 25-OCT-1995
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/328,519
FILING DATE: 25-OCT-1994
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Lin, Maria C.H.
REGISTRATION NUMBER: 29,323
REFERENCE/DOCKET NUMBER: 1151-4117
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-758-4800
TELEFAX: 212-751-6849
TELEX: 421792
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 25 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US95-13841-17

Query Match 6.4%; Score 124; DB 5; Length 25;
Best Local Similarity 84.0%; Pred. No. 1.6e-06;
Matches 21; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 103 ALNIWDRFDVFTLGASNGYIRGNS 127
|||||
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 7

US-08-446-692-24
Sequence 24, Application US/08446692
Patent No. 5759551
GENERAL INFORMATION:
APPLICANT: Ladd, Anna
APPLICANT: Wang, Chang Yi
APPLICANT: Zamb, Timothy
TITLE OF INVENTION: Immunogenic LHRH peptide constructs
and synthetic universal immune stimulators for vaccines
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Maria C.H. Lin
STREET: 345 Park Avenue
CITY: New York
STATE: NY
COUNTRY: US
ZIP: 10154-0053
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS

;; SOFTWARE: Patent In Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/446,692
;; FILING DATE: 7-JUN-1995
;; CLASSIFICATION: 424
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Maria C.H. Lin
;; REGISTRATION NUMBER: 29,323
;; REFERENCE/DOCKET NUMBER: 1151-4146 US2
;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212)415-8745
;; TELEFAX: (516)751-6849
;; INFORMATION FOR SEQ ID NO: 24:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 37 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-446-692-24

Query Match 6.4%; Score 124; DB 1; Length 37;
Best Local Similarity 84.0%; Pred. No. 3e-06; 1; Indels
Matches 21; Conservative 3; Mismatches 1; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 8
US-08-488-351A-24
; Sequence 24, Application US/08488351A
; Patent No. 5843446
; GENERAL INFORMATION:
; APPLICANT: Ladd, Anna
; APPLICANT: Wang, Chang Yi
; APPLICANT: Zamb, Timothy
; TITLE OF INVENTION: Immunogenic LHRH peptide constructs
; TITLE OF INVENTION: And synthetic universal immune stimulators for vaccines
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Maria C.H. Lin
; STREET: 345 Park Avenue
; CITY: New York
; STATE: NY
; COUNTRY: US
; ZIP: 10154-0053
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/488,351A
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/446,692
; FILING DATE: 7-JUN-1995
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/229,275
; FILING DATE: 14-APR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/057,166
; FILING DATE: 27-APR-1992
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Maria C.H. Lin
; REGISTRATION NUMBER: 29,323
; REFERENCE/DOCKET NUMBER: 1151-4146 US2

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212)415-8745
;; TELEFAX: (516)751-6849
;; INFORMATION FOR SEQ ID NO: 24:
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 37 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: peptide
US-08-488-351A-24

Query Match 6.4%; Score 124; DB 2; Length 37;
Best Local Similarity 84.0%; Pred. No. 3e-06;
Matches 21; Conservative 3; Mismatches 1; Indels
0; Gaps 0;

QY 103 ALNIWDRFDVFTLGASNGYIRGNS 127
Db 1 ALNIWDRFDVFTLGATGYLKGN 25

RESULT 9
US-08-519-385B-1
; Sequence 1, Application US/08519385B
; Patent No. 6001372
; GENERAL INFORMATION:
; APPLICANT: DeMars, Robert I.
; APPLICANT: Ortiz, Linette (n.m.i.)
; TITLE OF INVENTION: Antigenic Peptides of Chlamydia trachomatis
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Quarles & Brady
; STREET: 411 East Wisconsin Avenue
; CITY: Milwaukee
; STATE: WI
; COUNTRY: U.S.A.
; ZIP: 53202-4497
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/519,385B
; FILING DATE:
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Schwartz, Carl R.
; REGISTRATION NUMBER: 29,437
; REFERENCE/DOCKET NUMBER: 960296.93456
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (414) 277-5715
; TELEFAX: (414) 271-3552
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 31 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: no
; ANTI-SENSE: no
; FRAGMENT TYPE: no internal fragment
; ORIGINAL SOURCE:
; ORGANISM: Chlamydia trachomatis
; STRAIN: Serovar E
; INDIVIDUAL ISOLATE: MOMP Protein
US-08-519-385B-1

Query Match 6.1%; Score 119; DB 3; Length 31;
Best Local Similarity 64.5%; Pred. No. 7.5e-06;
Matches 20; Conservative 5; Mismatches 6; Indels
0; Gaps 0;

CLASSIFICATION: 530
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: P-40,403
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV7 - SOLV3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 834 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-471-033-21

Query Match 4.5%; Score 87.5; DB 1; Length 834;
 Best Local Similarity 20.2%; Pred. No. 2.8; Mismatches 122; Indels 133; Gaps 17;
 Matches 74; Conservative 38;
 Qy 46 DYVDFRLKVDAPKTFSGAKPTGSAANYTTAVDRPNPAYNKHLDHDAEFTNAGFIAL- 104
 Db 155 EYQSDTKFNIDS-KTFK-ELKLFKIDSQSQVOLRNPENKK-ESQEFLLAKASKTNLF 211
 Qy 105 -----NIWDRFDVFCILGASNGVIRGNSTAFNLVGLFGVKGTT-- 142
 Db 212 KQKMRDIDEDTDGDSIPDLWE-----ENGYTIQNKVAVKWDLSLASKGYTKF 261
 Qy 143 -----VNANEL-----PNVLSLN 155
 Db 262 VSNPLDSTHTGDPYTDYEKAARDLDSNAKETFNPLVAFFSVNVMKVIILSPNENLSN 321
 Qy 156 GVVELYDTSTFSWSVGARGALWECCATLGAEF-----OYAKSKPKVEELNVIC-NVSQFS 210
 Db 322 S-VESHSTNWSYT-NTEGASIEAGGGLGLSGFVSVTYQHSETVAQEWGTTGTGNTSQFN 379
 Qy 211 VNKPGYKGVAFPLPTDAGVATATGKATINYHEWQVGSLSYRLNSLVPYIGVQWSRA 270
 Db 380 -----TASAGYLNANRYN--NVGTGAIYDVKPTTSEFV----- 410
 Qy 271 TFDADNIRIAQPKLPTAVLNLTAWN-----PSLLGNATALSTTDSFSDF-MQIVSQCI 322
 Db 411 ---LNNNTIA---TITAKSNSTALRISPGDSYPEIGENAIATSMDDFNSHPITLNNKQOV 464
 Qy 323 NKFKSRK 329
 Db 465 NOLINNK 471

RESULT 15
 US-08-471-044-21
 ; Sequence 21, Application US/08471044
 ; Patent No. 5640868
 ; GENERAL INFORMATION:
 ; APPLICANT: Warren, Gregory W
 ; APPLICANT: Koziel, Michael G
 ; APPLICANT: Mullins, Martha A
 ; APPLICANT: Nye, Gordon J
 ; APPLICANT: Carr, Brian
 ; APPLICANT: Desai, Nalini M
 ; APPLICANT: Kostichka, N. Kristy
 ; APPLICANT: Duck, Nicholas B
 ; APPLICANT: Estruch, Juan J

TITLE OF INVENTION: No. 5840868el Pesticidal Proteins and Strains
 NUMBER OF SEQUENCES: 50
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: CIBA-GEIGY Corporation
 STREET: 7 Skyline Drive
 CITY: Hawthorne
 STATE: NY
 COUNTRY: USA
 ZIP: 10532
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30B
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/471,044
 FILING DATE: 06-JUN-1995
 CLASSIFICATION: 800
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/463,483
 FILING DATE: 05-JUN-1995
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/314,594
 FILING DATE: 09-SEP-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/218,018
 FILING DATE: 23-MAR-1994
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/037,057
 FILING DATE: 25-MAR-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Pace, Gary M.
 REGISTRATION NUMBER: 40,403
 REFERENCE/DOCKET NUMBER: CGC 1695/CIP3/DIV6 - SOLV3
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 919-541-8582
 TELEFAX: 919-541-8689
 INFORMATION FOR SEQ ID NO: 21:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 834 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-471-044-21

Query Match 4.5%; Score 87.5; DB 2; Length 834;
 Best Local Similarity 20.2%; Pred. No. 2.8; Mismatches 122; Indels 133; Gaps 17;
 Matches 74; Conservative 38;
 Qy 46 DYVDFRLKVDAPKTFSGAKPTGSAANYTTAVDRPNPAYNKHLDHDAEFTNAGFIAL- 104
 Db 155 EYQSDTKFNIDS-KTFK-ELKLFKIDSQSQVOLRNPENKK-ESQEFLLAKASKTNLF 211
 Qy 105 -----NIWDRFDVFCILGASNGVIRGNSTAFNLVGLFGVKGTT-- 142
 Db 212 KQKMRDIDEDTDGDSIPDLWE-----ENGYTIQNKVAVKWDLSLASKGYTKF 261
 Qy 143 -----VNANEL-----PNVLSLN 155
 Db 262 VSNPLDSTHTGDPYTDYEKAARDLDSNAKETFNPLVAFFSVNVMKVIILSPNENLSN 321
 Qy 156 GVVELYDTSTFSWSVGARGALWECCATLGAEF-----OYAKSKPKVEELNVIC-NVSQFS 210
 Db 322 S-VESHSTNWSYT-NTEGASIEAGGGLGLSGFVSVTYQHSETVAQEWGTTGTGNTSQFN 379
 Qy 211 VNKPGYKGVAFPLPTDAGVATATGKATINYHEWQVGSLSYRLNSLVPYIGVQWSRA 270
 Db 380 -----TASAGYLNANRYN--NVGTGAIYDVKPTTSEFV----- 410
 Qy 271 TFDADNIRIAQPKLPTAVLNLTAWN-----PSLLGNATALSTTDSFSDF-MQIVSQCI 322
 Db 411 ---LNNNTIA---TITAKSNSTALRISPGDSYPEIGENAIATSMDDFNSHPITLNNKQOV 464

Fri Feb 8 08:45:03 2002

us-09-391-606-16.rai

Page 8

QY 323 NKFKSRK 329
I: : |
Db 465 NOLINK 471

Search completed: February 7, 2002, 21:36:26
Job time: 20495 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:10 ; Search time 96.2 Seconds
(without alignments)
290.604 Million cell updates/sec

Title: US-09-391-606-16
Perfect score: 1952
Sequence: 1 MLPVGNPSPDPSLLIDGTWE.....TAEARLINERAAHVSGQFRF 367

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68:.*
1: pir1:.*
2: pir2:.*
3: pir3:.*
4: pir4:.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	99.7	389	2 A43587	major outer membra
2	1947	99.7	389	2 D86577	major outer membra
3	1910	97.8	389	2 I40864	major outer membra
4	1882	96.4	389	2 I40739	major outer membra
5	1534.5	78.6	389	1 MMCWP3	major outer membra
6	1496.5	76.7	389	2 A60109	major outer membra
7	1477	75.7	392	2 A40371	major outer membra
8	1419	72.7	402	2 I40740	major outer membra
9	1415	72.5	402	1 MMCWPM	major outer membra
10	1410	72.2	402	2 B60109	major outer membra
11	1408	72.1	402	2 A60341	major outer membra
12	1351.5	69.2	387	2 C81747	major outer membra
13	1351	69.2	404	2 I40741	major outer membra
14	1349.5	69.1	387	2 J70947	major outer membra
15	1349.5	69.1	387	2 S16034	mouse pneumonitis
16	1340.5	68.7	397	1 MMCWTH	major outer membra
17	1338.5	68.6	397	2 JF0413	major outer membra
18	1333.5	68.3	375	2 S11007	major outer membra
19	1328	68.0	374	2 S11006	major outer membra
20	1325	67.9	396	2 S12799	major outer membra
21	1315.5	67.4	393	1 MMCWTE	major outer membra
22	1315.5	67.4	397	1 MMCWTC	major outer membra
23	1313.5	67.3	393	2 S06259	major outer membra
24	1310.5	67.1	393	2 T01645	major outer membra
25	1308.5	67.0	393	2 H71484	probable major out
26	1307.5	67.0	395	1 MMCWTF	major outer membra
27	1303.5	66.8	393	2 JC1432	major outer membra
28	1295	66.3	394	2 S11012	major outer membra
29	1292	66.2	372	2 B60756	major outer membra

30	1285	65.8	372	2 S11009	major outer membra
31	1275	65.3	394	1 MMCWTB	major outer membra
32	373	19.1	158	2 A60333	outer membrane pro
33	361	18.5	156	2 B60333	outer membrane pro
34	355.5	18.2	155	2 C60333	outer membrane pro
35	215.5	11.0	136	2 F30593	outer membrane pro
36	215	11.0	135	2 C30593	outer membrane pro
37	212.5	10.9	136	2 D30593	outer membrane pro
38	206.5	10.7	136	2 B30593	outer membrane pro
39	206.5	10.6	136	2 I30587	outer membrane pro
40	206.5	10.6	136	2 E30593	outer membrane pro
41	203.5	10.4	136	2 A30593	outer membrane pro
42	189	9.7	340	2 H71479	probable outer mem
43	186.5	9.7	134	2 G30587	outer membrane pro
44	182	9.3	340	2 H81742	major outer membra
45	179.5	9.2	134	2 H30587	outer membrane pro

ALIGNMENTS

RESULT 1

A43587

major outer membrane protein, porin CP0051 precursor [imported] - Chlamydoiphila pneu
N:Alternate names: MOMP
C:Species: Chlamydoiphila pneumoniae, Chlamydia pneumoniae
C:Date: 29-Jan-1993 #sequence revision 29-Jan-1993 #text_change 11-May-2000
C:Accession: A43587; A49751; A49216; G72044; F81619
R:Perez Melgosa, M.; Kuo, C.C.; Campbell, L.A.
Infect. Immun. 59, 2195-2199, 1991
A:Title: Sequence analysis of the major outer membrane protein gene of Chlamydia pne
A:Reference number: A43587; MUID:91244474
A:Accession: A43587
A:Molecule type: DNA
A:Residues: 1-389 <PER>
A:Cross-references: GB:M69230; NID:g144540; PIDN:AAA73071.1; PID:g144541
R:Carter, M.W.; Al-Mahdawi, S.A.H.; Giles, I.G.; Treharne, J.D.; Ward, M.E.; Clarke, J. Gen. Microbiol. 137, 465-475, 1991
A:Title: Nucleotide sequence and taxonomic value of the major outer membrane protein
A:Reference number: A49751; MUID:91237311
A:Accession: A49751
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <CAR>
A:Cross-references: GB:M64064; GB:M34942; NID:g144534; PIDN:AAA23143.1; PID:g144535
A:Note: isolate IOL-207
R:Gaydos, C.A.; Quinn, T.C.; Bobo, L.D.; Eiden, J.J.
Infect. Immun. 60, 5319-5323, 1992
A:Title: Similarity of Chlamydia pneumoniae strains in the variable domain IV region
A:Reference number: A49216; MUID:93084388
A:Accession: A49216
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 297-352 <GAV>
A:Cross-references: GB:S50607; NID:g260972; PIDN:AAB24363.1; PID:g260973
A:Note: sequence extracted from NCBI backbone (NCBIN:120604, NCBI:P:120605)
R:Kalan, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Ollinger, L.; Grimwood, N. Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: G72044
A:Molecule type: DNA
A:Residues: 1-389 <ARN>
A:Cross-references: GB:AB001652; GB:AB001363; NID:g4376997; PIDN:AAD18834.1; PID:g4376997
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hick, C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg, S.L. Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPh and Chlamydia pneumoniae AR3
A:Reference number: A81500; MUID:20150255
A:Accession: F81619
A>Status: preliminary
A:Molecule type: DNA

A:Residues: 1-389 <REA>
A:Cross-references: GB:AE002168; GB:AE002161; NID:97188982; PIDN:AAF37944.1; PID:9718899
A:Experimental source: strain AR39, HL cells
C:Genetics:

A:Gene: ompA; CP0051
A:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F:1-23/Domain: signal sequence #status predicted <SIG>
F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 99.7%; Score 1947; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.2e-150;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 2 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
      |||
Db 24 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
      |||

QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCILGASNG 121
      |||
Db 84 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCILGASNG 143
      |||

QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTSFSWSVGARGALWECGC 181
      |||
Db 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTSFSWSVGARGALWECGC 203
      |||

QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKTSATI 241
      |||
Db 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKTSATI 263
      |||

QY 242 NYHEQVQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
      |||
Db 264 NYHEQVQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
      |||

QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVLDADKWSLTAEARLINERAHV 361
      |||
Db 324 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVLDADKWSLTAEARLINERAHV 383
      |||

QY 362 SQQFRF 367
      |||||
Db 384 SQQFRF 389
      |||||
```

RESULT 2
D86577
major outer membrane protein [imported] - Chlamydia pneumoniae (strain J138)
C:Species: Chlamydia pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: D86577
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Is
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: D86577
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-389 <STC>
A:Cross-references: GB:BA000008; NID:98979067; PIDN:BA989802.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:

A:Gene: ompA
C:Superfamily: Chlamydia major outer membrane protein

Query Match 99.7%; Score 1947; DB 2; Length 389;
Best Local Similarity 100.0%; Pred. No. 2.2e-150;
Matches 366; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 2 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
      |||
Db 24 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
      |||
```

```
QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCILGASNG 121
      |||
Db 84 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCILGASNG 143
      |||

QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTSFSWSVGARGALWECGC 181
      |||
Db 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTSFSWSVGARGALWECGC 203
      |||

QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKTSATI 241
      |||
Db 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKTSATI 263
      |||

QY 242 NYHEQVQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
      |||
Db 264 NYHEQVQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
      |||

QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVLDADKWSLTAEARLINERAHV 361
      |||
Db 324 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVLDADKWSLTAEARLINERAHV 383
      |||

QY 362 SQQFRF 367
      |||||
Db 384 SQQFRF 389
      |||||
```

RESULT 3
I40864

major outer membrane protein - Chlamydia psittaci
C:Species: Chlamydia psittaci, Chlamydia psittaci
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
C:Accession: I40864; S33465
R:Garjes, A.A.; Carrick, F.N.; Lavin, M.F.
Gene 138, 139-142, 1994
A:Title: Remarkable sequence relatedness in the DNA encoding the major outer membra
A:Reference number: I40864; MUID:94171025
A:Accession: I40864
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-389 <RES>
A:Cross-references: EMBL:X72023; NID:g313844; PIDN:CAA50906.1; PID:g313845
C:Superfamily: Chlamydia major outer membrane protein

Query Match 97.8%; Score 1910; DB 2; Length 389;
Best Local Similarity 97.5%; Pred. No. 2.2e-147;
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

```
QY 2 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
      |||
Db 24 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
      |||

QY 62 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCILGASNG 121
      |||
Db 84 SMGAKPTGSAANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCILGASNG 143
      |||

QY 122 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTSFSWSVGARGALWECGC 181
      |||
Db 144 YIRGNSTAFNLVGLFGVKGTTVNANELPNVSLNSGVVVELYDTSFSWSVGARGALWECGC 203
      |||

QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKTSATI 241
      |||
Db 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGKTSATI 263
      |||

QY 242 NYHEQVQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 301
      |||
Db 264 NYHEQVQASLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVNLNTAWNPSLLGN 323
      |||

QY 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTGATLVLDADKWSLTAEARLINERAHV 361
      |||
Db 324 TTTATSDSFDMQIVSCQINKFKSRKACGVTGATLVLDADKWSLTAEARLINERAHV 383
      |||

QY 362 SQQFRF 367
      |||||
```

Db 384 SQQFRF 389

RESULT 4

I40739

major outer membrane protein precursor - Chlamydomophila pneumoniae (strain equine/N16)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

A:Variety: strain equine/N16

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 20-Apr-2000

C:Accession: I40739

R:Storey, C.; Lusher, M.; Yates, P.; Richmond, S.

J. Gen. Microbiol. 139, 2621-2626, 1993

A:Title: Evidence for Chlamydia pneumoniae of non-human origin.

A:Reference number: I40739

A:Accession: I40739

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-389 <STO>

C:Cross-references: GB:I40982; NID:g289840; PIDN:AAAL17397.1; PID:g289841

C:Comment: On the basis of the major outer membrane protein the authors classified the

the sequence of the genome strain CWL029 and strain IOL-207. See PIR:A43587.

C:Genetics:

A:Gene: momp

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

F:1-23/Domain: signal sequence #status predicted <SIG>

F:24-389/Product: major outer membrane protein #status predicted <MAT>

Query Match

Best Local Similarity 96.4%; Score 1882; DB 2; Length 389;

Matches 346; Conservative 15; Mismatches 5; Indels 0; Gaps 0;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61

Db 24 LPVGNPDPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 83

QY 62 SNGAKPTGSAANNTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRDFVCTLGASNG 121

Db 84 SNGAKPTGSAATNTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRDFVCTLGASNG 143

QY 122 YIRGNSTAFNLVGLFGVKTGTTNANLNPVLSNGVVELYTDTSFSWSVGARGALWECG 181

Db 144 YKNGSAAFNVLGFGVKTGTTNANLNPVLSNGVVELYTDTSFSWSVGARGALWECG 203

QY 182 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 241

Db 204 ATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 263

QY 242 NYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 301

Db 264 NYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 323

QY 302 ATALSTTDSFDMQIVSQINPKFKSRKACGTVGATLVDAKWSLTAEARLINERAHV 361

Db 324 ATAVSSDDQFSDQMIVSQINPKFKSRKACGTVGATLVDAKWSLTAEARLINERAHV 383

QY 362 SQQFRF 367

Db 384 SQQFRF 389

RESULT 5

MNCWP3

major outer membrane protein precursor - Chlamydomophila psittaci (strain S26/3)

C:Species: Chlamydomophila psittaci, Chlamydia psittaci

C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 31-Mar-2000

C:Accession: S08770

R:Herring, A.J.; Tan, T.W.; Baxter, S.; Inglis, N.F.; Dunbar, S.

FEMS Microbiol. Lett. 65, 153-158, 1989

A:Title: Sequence analysis of the major outer membrane protein gene of an ovine abortion

A:Reference number: S08770

A:Accession: S08770

A:Molecule type: DNA

A:Residues: 1-389 <HER>

A:Cross-references: EMBL:X51859; NID:g40600; PIDN:CAA36152.1; PID:g40601

C:Superfamily: Chlamydia major outer membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match

Best Local Similarity 78.6%; Score 1534.5; DB 1; Length 389;

Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61

Db 23 LPVGNPAEPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDVKNKI 82

QY 62 S-MGAKPTGSAANNTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRDFVCTLGASN 120

Db 83 TCMGAVPTGTAAANTKTPTDRPNIAYGKHLQDAEFTNAAFLALNIWDRDFICTLGASN 142

QY 121 GYIRGNSTAFNLVGLFGVKTGTTNANLNPVLSNGVVELYTDTSFSWSVGARGALWECG 180

Db 143 GYFKASSRAAFNLVGLIGVKGSSIAADOLPNVGITGIVEFYDDTTFSSVGARGALWECG 202

QY 181 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 240

Db 203 CATLGAEFQYAOQSKPKVEELNVCNVSQFSVKNPKGYKGVAFPLPTDAGVATATGTSAT 262

QY 241 INYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 300

Db 263 IYHEWQVGLSRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 322

QY 301 NATALSTTDSFDMQIVSQINPKFKSRKACGTVGATLVDAKWSLTAEARLINERAHV 360

Db 323 EATALDTSNKFADFQIASIQINKMKSRKACGTVGATLVDAKWSLTAEARLINERAHV 382

QY 361 VSGQFRF 367

Db 383 MNAQFRF 389

RESULT 6

A60109

major outer membrane protein precursor - Chlamydomophila psittaci (strain Guinea pig 1

C:Species: Chlamydomophila psittaci, Chlamydia psittaci

C:Date: 10-Nov-1992 #sequence_revision 10-Nov-1992 #text_change 31-Mar-2000

C:Accession: A60109

R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.

Infect. Immun. 57, 1621-1625, 1989

A:Title: Cloning and sequence analysis of the major outer membrane protein genes of

A:Reference number: A60109; MUID:89212917

A:Accession: A60109

A:Status: not compared with conceptual translation

A:Molecule type: DNA

A:Residues: 1-389 <ZHA>

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

F:1-22/Domain: signal sequence #status predicted <SIG>

F:23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match

Best Local Similarity 76.7%; Score 1496.5; DB 2; Length 389;

Matches 272; Conservative 42; Mismatches 51; Indels 3; Gaps 2;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61

Db 23 LPVGNPAEPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDVKNKI 82

QY 62 SNGAKPTGSAANNTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRDFVCTLGASNG 121

Db 83 SNGTAPTGNAAADFXTVADRNNIAYGKHKMQDAEFTNAAFLALNIWDRDFVCTLGASNG 142

QY 122 YIRGNSTAFNLGFGVKGTTNNANLPVLSNGWVLYTDTFSWSVGARGALWECG 181
Db 143 YLKANAFAFNLGLGVTGTDQ-QGYPNAISQGLVELYDTTFSWSVGARGALWECG 201
QY 182 ATLGAEFQYQAQSKPKVEELNVCNVSQFVNPKPKYGVA--FPLPTDAGVATATGTXA 239
Db 202 ATLGAEFQYQAQSKPKVEELNVCNVSQFVNPKPKYGVA--FPLPTDAGVATATGTXA 261
QY 240 TTYNHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTANNPSIL 299
Db 262 TTYNHEWQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTANNPSIL 321
QY 300 GNATALSTTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINERAA 359
Db 322 GNATALSTTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINERAA 381
QY 360 HVSGQFRF 367
Db 382 HVNAQFRF 389
RESULT 7
A40371
major outer membrane protein precursor - Chlamydomophila psittaci (strain Fpn/pring)
C:Species: Chlamydomophila psittaci, Chlamydia psittaci
C:Date: 27-Nov-1991 #sequence_revision 27-Nov-1991 #text_change 31-Mar-2000
C:Accession: A40371; S16137
J. Storey, C.; Lusher, M.; Yates, P.; Richmond, S.
R:Gen. Microbiol. 139, 2621-2626, 1993
A:Title: Evidence for Chlamydia pneumoniae of non-human origin.
A:Reference number: 140739; MUID:94103736
A:Accession: 140859
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB/EMBL
A:Residues: 1-392 <RES>
A:Cross-references: EMBL:X61096; NID:940564; PIDN:CAA43409.1; PID:940565
A:Experimental source: strain Fpn
C:Genetics:
A:Gene: MOMP
C:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-392/Product: major outer membrane protein #status predicted <MAT>
Query Match 75.7%; Score 1477; DB 2; Length 392;
Best Local Similarity 74.1%; Pred. No. 3e-112;
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;
QY 2 LPVGNPSPSLIDGTIWEGAAGDPCDPCATWCDALISIRAGFYGYVDFRILKVDPAKTF 61
Db 23 LPVGNPAEPSLLIDGTWEGASGDPCDPCATWCDALISIRAGFYGYVDFRILKVDVKNKI 82
QY 62 S-MGAKPTGSA--AANYTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDVECTIGA 118
Db 83 SCMAAPTAASGTATNVAADRSFAYGKHLDHAEFTNAGFIALNIWDRDVECTIGA 142
QY 119 SNGYIRGNSTAFNLGFGVKGTTNNANLPVLSNGWVLYTDTFSWSVGARGALWE 178
Db 143 SNGYIRGNSTAFNLGFGVKGTTNNANLPVLSNGWVLYTDTFSWSVGARGALWE 201
QY 179 CGCATLGAEFQYQAQSKPKVEELNVCNVSQFVNPKPKYGVA--FPLPTDAGVATATGT 236
Db 202 CGCATLGAEFQYQAQSKPKVEELNVCNVSQFVNPKPKYGVA--FPLPTDAGVATATGT 261
QY 237 KSATNHYEMQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTANNP 296
Db 262 KSATNHYEMQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTANNP 321
QY 297 SLIGNATALSTTDSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLINE 356
Db 322 TLLGVATTLTDSNKNYADFMQIVSMQINKFKSRKACGVTVGATLVADKWSLTAEARLINE 381

QY 357 RAAHVSGQFRF 367
Db 382 RAAHINAQFRF 392
RESULT 8
I40740
major outer membrane protein - Chlamydomophila psittaci (strain N352)
C:Species: Chlamydomophila psittaci, Chlamydia psittaci
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
C:Accession: I40740
J. Storey, C.; Lusher, M.; Yates, P.; Richmond, S.
R:Gen. Microbiol. 139, 2621-2626, 1993
A:Title: Evidence for Chlamydia pneumoniae of non-human origin.
A:Reference number: 140739; MUID:94103736
A:Accession: I40740
A:Status: nucleic acid sequence not shown; translation not shown; translated from GB
A:Molecule type: DNA
A:Residues: 1-402 <RES>
A:Cross-references: GB:L04980; NID:9144544; PIDN:AAA17396.1; PID:9144545
C:Genetics:
A:Gene: momp
C:Superfamily: Chlamydia major outer membrane protein
Query Match 72.7%; Score 1419; DB 2; Length 402;
Best Local Similarity 70.5%; Pred. No. 1.6e-107;
Matches 268; Conservative 37; Mismatches 61; Indels 14; Gaps 6;
QY 2 LPVGNPSPSLIDGTIWEGAAGDPCDPCATWCDALISIRAGFYGYVDFRILKVDPAKTF 61
Db 23 LPVGNPAEPSLLIDGTWEGASGDPCDPCATWCDALISIRAGFYGYVDFRILKVDVKNKI 82
QY 62 S-MGAKP---TCSAA-ANYTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRDVECTIL 116
Db 83 SCMAAPTAASGTATNVAADRSFAYGKHLDHAEFTNAGFIALNIWDRDVECTIL 142
QY 117 GASNGYIRGNSTAFNLGFGVKGTTNNANLPVLSNGWVLYTDTFSWSVG 172
Db 143 GASNGYIRGNSTAFNLGFGVKGTTNNANLPVLSNGWVLYTDTFSWSVG 202
QY 173 RGALECCATLGAEFQYQAQSKPKVEELNVCNVSQFVNPKPKYGVA--FPLPTDAGV 230
Db 203 RGALECCATLGAEFQYQAQSKPKVEELNVCNVSQFVNPKPKYGVA--FPLPTDAGV 262
QY 231 ATATGTSKATNHYEMQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
Db 263 ATATGTSKATNHYEMQVGSLSYRLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLN 322
QY 291 LTANNPSLLGNATALSTT---DSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSL 347
Db 323 LTANNPSLLGNATALSTT---DSFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSL 382
QY 348 TAEARLINERAAHVSGQFRF 367
Db 383 TAEARLINERAAHVSGQFRF 402
RESULT 9
MNCWPM
major outer membrane protein precursor - Chlamydomophila psittaci (strain A22/M)
C:Species: Chlamydomophila psittaci, Chlamydia psittaci
C:Date: 31-Dec-1990 #sequence_revision 31-Dec-1990 #text_change 31-Mar-2000
C:Accession: S05954
R:Pickett, M.A.; Everson, J.S.; Clarke, I.N.
FEMS Microbiol. Lett. 55, 229-234, 1988
A:Title: Chlamydia psittaci ewe abortion agent: complete nucleotide sequence of the
A:Reference number: S05954
A:Accession: S05954
A:Molecule type: DNA
A:Residues: 1-402 <PIC>
A:Cross-references: EMBL:X12647; NID:940604; PIDN:CAA31177.1; PID:940605
C:Superfamily: Chlamydia major outer membrane protein

F.1-22/Domain: signal sequence #status predicted <SIG>
F.23-402/Product: major outer membrane protein #status predicted <MAT>

Query Match 72.5%; Score 1415; DB 1; Length 402;
Best Local Similarity 70.3%; Pred. No. 3.4e-107;
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTWEGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSSLIDGTWEGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSA--ANTYTTAVDRPNPAYNKHLDHAEFTNAGFTALNINWDFVFCFL 116
DB 83 SGMAATPTQATGNASNTNPEANGRNIAIYGRHMDAEWFSNAAFALNALINWDFDIFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG---LPNVSLSGVVELYTDTSFWSVGA 172
DB 143 GASNGYFKSSAFAFNVLGLIGFSSTSTELPMQLPNVIGITQGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSOFSVKNPKGYKGA--FPLPTDAGV 230
DB 203 RGALWECGCATLGAEFQYQAQSNPKIEVLNVTSSPAQFVHKPRGYKGASSNEPLPITACT 262

QY 231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPTAVLN 290
DB 263 TEATDTKSAITIKYHEWQVGLSYRLNMLVPYIGVWNSRATEDADTIRIAQPKLASEILN 322

QY 291 LTAWNPSSLGNATALSTT---DSFDFMOIVSCQINKFKSRKACGVTVGATLVADKWSL 347
DB 323 ITTWNPSLLGSTTTLPNNNGKDVLSVLQIASIQINKMKSRKACGVAVGATLIDADKWSI 382

QY 348 TAEARLINERAHVSGQFRF 367
DB 383 TGEARLINERAHMAQFRF 402

RESULT 10
B60109
major outer membrane protein precursor - Chlamydia psittaci (strain meningopneumoni
C:Species: Chlamydia psittaci, Chlamydia psittaci
C:Date: 10-Nov-1992 #sequence_revision 10-Nov-1992 #text_change 31-Mar-2000
C:Accession: B60109
R:Zhang, Y.X.; Morrison, S.G.; Caldwell, H.D.; Baehr, W.
Infected. Immun. 57, 1621-1625, 1989
A:Title: Cloning and sequence analysis of the major outer membrane protein genes of two
A:Reference number: A60109; MUID:89212917
A:Accession: B60109
A>Status: not compared with conceptual translation
A:Molecule type: DNA
A:Residues: 1-402 <ZNA>
C:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F.1-22/Domain: signal sequence #status predicted <SIG>
F.23-389/Product: major outer membrane protein #status predicted <MAT>

Query Match 72.2%; Score 1410; DB 2; Length 402;
Best Local Similarity 70.3%; Pred. No. 8.6e-107;
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTWEGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSSLIDGTWEGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSA--ANTYTTAVDRPNPAYNKHLDHAEFTNAGFTALNINWDFVFCFL 116
DB 83 SGMAATPTQATGNASNTNPEANGRNIAIYGRHMDAEWFSNAAFALNALINWDFDIFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVANLPLNVSLSGVVELYTDTSFWSVGA 172
DB 143 GASNGYFKSSAFAFNVLGLIGFSAASSISTDLPTQLPNVIGITQGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSOFSVKNPKGYKGA--FPLPTDAGV 230
DB 203 RGALWECGCATLGAEFQYQAQSNPKIEVLNVTSSPAQFVHKPRGYKGASSNEPLPITACT 262

QY 231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPTAVLN 290
DB 263 TEATDTKSAITIKYHEWQVGLSYRLNMLVPYIGVWNSRATEDADTIRIAQPKLASEILN 322

QY 291 LTAWNPSSLGNATAL---STTDSFDFMOIVSCQINKFKSRKACGVTVGATLVADKWSL 347
DB 323 ITTWNPSLLGSTTTLPNNNGKDVLSVLQIASIQINKMKSRKACGVAVGATLIDADKWSI 382

QY 348 TAEARLINERAHVSGQFRF 367
DB 383 TGEARLINERAHMAQFRF 402

RESULT 11
A60341
major outer membrane protein precursor - Chlamydia psittaci (strain 6BC)
C:Species: Chlamydia psittaci, Chlamydia psittaci
C:Date: 11-Dec-1992 #sequence_revision 24-Feb-1994 #text_change 31-Mar-2000
C:Accession: A44565; A60341; B60341
R:Everett, K.D.E.
submitted to the EMBL Data Library, December 1990
A:Reference number: A44565
A:Accession: A44565
A:Molecule type: DNA
A:Residues: 1-402 <EV>
A:Cross-references: GB:X56980; NID:g40568; PIDN:CAA40300.1; PID:g40569
R:Everett, K.D.E.; Andersen, A.A.; Plaunt, M.; Hatch, T.P.
Infected. Immun. 59, 2853-2855, 1991
A:Title: Cloning and sequence analysis of the major outer membrane protein gene of
A:Reference number: A60341; MUID:91310346
A:Accession: A60341
A:Molecule type: protein
A:Residues: 23-35 <EV2>
A:Accession: B60341
A>Status: nucleic acid sequence not shown
A:Molecule type: DNA
A:Residues: 112-232; 317-349 <EV3>
A:Cross-references: GB:X56980
C:Superfamily: Chlamydia major outer membrane protein
C:Keywords: membrane protein
F.1-22/Domain: signal sequence #status predicted <SIG>
F.23-402/Product: major outer membrane protein #status experimental <MAT>

Query Match 72.1%; Score 1408; DB 2; Length 402;
Best Local Similarity 70.3%; Pred. No. 1.2e-106;
Matches 267; Conservative 38; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTWEGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSSLIDGTWEGAGDPCDPCATWCDALISIRAGYGYGVDFRILKVDVNTKF 82

QY 62 S-MGAKP---TGSA--ANTYTTAVDRPNPAYNKHLDHAEFTNAGFTALNINWDFVFCFL 116
DB 83 SGMAATPTQATGNASNTNPEANGRNIAIYGRHMDAEWFSNAAFALNALINWDFDIFCTL 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKG---TTVANLPLNVSLSGVVELYTDTSFWSVGA 172
DB 143 GASNGYFKSSAFAFNVLGLIGFSAASSISTDLPMQLPNVIGITQGVVEFTDTSFWSVGA 202

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSOFSVKNPKGYKGA--FPLPTDAGV 230
DB 203 RGALWECGCATLGAEFQYQAQSNPKIEVLNVTSSPAQFVHKPRGYKGASSNEPLPITACT 262

QY 231 AFATGKTSATINHEWQVGSLSYRLNSLVPYIGVOWSRATEDADNIRIAQPKLPTAVLN 290
DB 263 TEATDTKSAITIKYHEWQVGLSYRLNMLVPYIGVWNSRATEDADTIRIAQPKLASEILN 322

QY 291 LTAWNPSSLGNATAL---STTDSFDFMOIVSCQINKFKSRKACGVTVGATLVADKWSL 347

Db 323 ITTNPSLIGSTALPNNSGKDVLSVLQIASIQINKMSRKACGAVAGATLIDAKWSI 382
QY 348 TAEARLINERAAHVSGQFRF 367
Db 383 TGEARLINERAAHVNAQFRF 402

RESULT 12

CB1747

major outer membrane protein, porin TC0052 [imported] - Chlamydia muridarum (strain Nigg)

C:Species: Chlamydia muridarum, Chlamydia trachomatis MoPn

C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-May-2000

C:Accession: CB1747

R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,

C.; Dodson, R.; Gwin, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,

Nucleic Acids Res. 28, 1397-1406, 2000

A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.

A:Reference number: AB1500; MUID:20150255

A:Accession: CB1747

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-387 <RES>

A:Cross-references: GB:AB002272; GB:AB002160; NID:g7190083; PIDN:AAF38941.1; PID:g719009

A:Experimental source: strain Nigg (MoPn)

C:Genetics:

A:Gene: TC0052

C:Superfamily: Chlamydia major outer membrane protein

Query Match 69.2%; Score 1351.5; DB 2; Length 387;
Best Local Similarity 66.4%; Pred. No. 4.5e-102;
Matches 245; Conservative 50; Mismatches 67; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLLDGTIWEAGAGPCPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61
Db 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISLRGLGYGVDFVDRVLKTDVNVKQF 82
QY 62 SMGAKPTGSAANYTTA---VDPRNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFCITLGA 118
Db 83 EMGAFTGD--ADLTATTPASRENPAYGKHQDAEMFTNAAYMALNIWDRFDVFCITLGA 140
QY 119 SNGYIRGNSTAFNLVGLFGVKGTNNANLPVNSLNGVVELYDTFSWSVGARGALWE 178
Db 141 TSGYLKGNASAFNLVGLFRDTEAADDIPNVLSQAQVVELYDTAFASVSGARALWE 200
QY 179 CCATILGAEFOYAQSKPKVEELNVICNVSFQSVNPKGYKGVAFPLPTDAGVATATGTS 238
Db 201 CGCATILGASFOYAQSKPKVEELNVLCNAAEFTINKPKGYGQEFPLNKAGTVSATDTKD 260
QY 239 ATINHEWQVGLASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSL 298
Db 261 ASIDHEWQASLALSYRLNMFTPIYGVKWSRASFDADTIRIAQPKLETISILMTWNTPTI 320
QY 299 LGNATALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLITAEARLINERA 358
Db 321 SSGSIDVDLT--KITDTLQIVSLQNLKMSRKSCGLAIGTTIVADVADKYAVTVETRLIDERA 378
QY 359 AHVSGQFRF 367
Db 379 AHVNAQFRF 387

RESULT 13

I40741

major outer membrane protein - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 20-Aug-1999

C:Accession: I40741

R:Zhang, Y.X.; Fox, J.G.; Ho, Y.; Zhang, L.; Stills, H.F.; Smith, T.F.

Mol. Biol. Evol. 10, 1327-1342, 1993

A:Title: Comparison of the major outer-membrane protein (MOMP) gene of mouse pneumonitis

A:Reference number: A49379; MUID:94104488

A:Accession: I40741
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-404 <RES>
A:Cross-references: GB:L19221; NID:g410146; PIDN:AAA16615.1; PID:g410147
C:Genetics:
A:Gene: momp
C:Superfamily: Chlamydia major outer membrane protein

Query Match 69.2%; Score 1351; DB 2; Length 404;
Best Local Similarity 65.0%; Pred. No. 5.3e-102;
Matches 249; Conservative 50; Mismatches 66; Indels 18; Gaps 4;

QY 2 LPVGNPSPSLLDGTIWEAGAGPCPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61
Db 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISLRGLGYGVDFVDRVLKTDVNVKQF 82
QY 62 SMGAKPTGSG---AAANYTTAVDRNPAYNKHLDHAEWFTNAGFIALNIWDRFDVFCITLGA 118
Db 83 EMGPVPTTDDTDAADITSTPRENPAYGKHQDAEMFTNAAYMALNIWDRFDVFCITLGA 142
QY 119 SNGYIRGNSTAFNLVGLFGVKG-----TTVNANLPVNSLNGVVELYDTFSWSV 170
Db 143 TSGYLKGNASAFNLVGLFG-DCVANANAIAATVAADSLPNVLSQAQVVELYDTAFANSV 201
QY 171 GARGALWECGCATILGAEFOYAQSKPKVEELNVICNVSFQSVNPKGYKGVAFPLPTDAGV 230
Db 202 GARALWECGCATILGASFOYAQSKPKVEELNVLCNAOFTINKPKGYGKEFFPLALTAGT 261
QY 231 ATATGKTGATINHEWQVGLASLYRLNSLPYIGVQWSRATFDADNIRIAQPKLPTAVLN 290
Db 262 DSATDTKASIDYHEWQASLALSYRLNMFTPIYGVKWSRASFDADTIRIAQPKLAFILD 321
QY 291 LTANNPSLLGNAT-----ALSTDSFDMQIVSCQINKFKSRKACGVTVGATLVADK 344
Db 322 VTTWNPTTAGAGTTADGTGAAATANGADTLQIVSLQNLKMSRKSCGLAIGTTIVADK 381
QY 345 WSLTAEARLINERAAHVSGQFRF 367
Db 382 YAVTVETRLIDERAHVNAQFRF 404

RESULT 14

JT0947

mouse pneumonitis major outer membrane protein - Chlamydia trachomatis

C:Species: Chlamydia trachomatis

C:Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 20-Aug-1999

C:Accession: JT0947

R:Fielder, T.J.; Pal, S.; Peterson, E.M.; de la Maza, L.M.

Gene 106, 137-138, 1991

A:Title: Sequence of the gene encoding the major outer membrane protein of the mouse

A:Reference number: JT0947; MUID:92039057

A:Accession: JT0947

A:Molecule type: DNA

A:Residues: 1-387 <FE>

A:Cross-references: GB:M64171; NID:g144536; PIDN:AAA23144.1; PID:g144537

C:Comment: C. trachomatis are gram negative bacteria that cause a variety of infections

C:Genetics:

A:Gene: om1MoPn

C:Superfamily: Chlamydia major outer membrane protein

C:Keywords: membrane protein

Query Match 69.1%; Score 1349.5; DB 2; Length 387;
Best Local Similarity 66.1%; Pred. No. 6.6e-102;
Matches 244; Conservative 50; Mismatches 68; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLLDGTIWEAGAGPCPCATWCDAISLRAGFYGYVDFRILKVDAPKTF 61
Db 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISLRGLGYGVDFVDRVLKTDVNVKQF 82

QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEFTNAGFIALNIWDRFDVFTLGA 118
 DB 83 EMGAAPTGD--ADLTPTAPASRENPAKQKQDAEMFTNAAVMALNIWDRFDVFTLGA 140
 QY 119 SNGYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWE 178
 DB 141 TSGYLKGNAAFNVLGLFGRDETAAADDIPNVLSQLAVVELYTDTSFWSVGARGALWE 200
 QY 179 CCATLGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKGYKGVAFPLPTDAGVATATGKS 238
 DB 201 CGCPTLGASFOYQAQSKPKVEELNVICNAAEFTINPKGVGQEFPLNKAGTVSATDTKD 260
 QY 239 ATINHEWQVGSLSYRLNSLPYIGVQWRSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 298
 DB 261 ASIDYHEWQASLSYRLNMFTPIYIGVQWRSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 320
 QY 299 LGNATALSTTDSFSDFMQIVSCQINKFKSRKACGTVGATLVADKWSLTAEARLINERA 358
 DB 321 SGSGIDVDT--KITDTLQIVSLQNLNKKSRKSCGLAIGTTIVDADKYAVTVETRLIDERA 378
 QY 359 AHVSGQFRF 367
 DB 379 AHVNAQFRF 387

RESULT 15

SI6034
 major outer membrane protein - Chlamydia trachomatis
 C:Species: Chlamydia trachomatis
 C>Date: 20-Feb-1995 #sequence_revision 20-Feb-1995 #text_change 12-Apr-1996
 C:Accession: SI6034; A43875
 R:Zhang, Y.X.; Zhang, L.
 submitted to the EMBL Data Library, July 1991
 A:Description: The nucleotide sequence of major outer membrane protein gene of mouse bio
 A:Reference number: SI6034
 A:Accession: SI6034
 A>Status: Preliminary
 A:Molecule type: DNA
 A:Residues: 1-387 <ZHA>
 R:Peterson, E.M.; Cheng, X.; Markoff, B.A.; Fielder, T.J.; de la Maza, L.M.
 Infect. Immun. 59, 4147-4153, 1991
 A:Title: Functional and structural mapping of Chlamydia trachomatis species-specific ma
 A:Reference number: A43875; MUID:92040090
 A:Accession: A43875
 A>Status: Preliminary; not compared with conceptual translation
 A:Molecule type: nucleic acid
 A:Residues: 314-322 <PET>
 A:Experimental source: mouse pneumonitis strain
 A>Note: sequence extracted from NCBI backbone (NCBIP:62877)
 C:Superfamily: Chlamydia major outer membrane protein

Query Match 69.1%; Score 1349.5; DB 2; Length 387;
 Best Local Similarity 66.1%; Pred No. 6,6e-102;
 Matches 244; Conservative 50; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPDLLIDTIFWGAAGPCDPCATWCDALSRAGFYGDYFDRILKVDAPKTF 61
 DB 23 LPVGNPAEPLMIDGLTWEGFGDPCDCTTWCDAISLRGLYGGDFVDRVLKTDVKNQF 82
 QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEFTNAGFIALNIWDRFDVFTLGA 118
 DB 83 EMGAAPTGD--RDLTPTAPASRENPAKQKQDAEMFTNAAVMALNIWDRFDVFTLGA 140
 QY 119 SNGYIRGNSTAFNLVGLFGVKTNNANLNPVLSNGVVELYTDTSFWSVGARGALWE 178
 DB 141 TSGYLKGNAAFNVLGLFGRDETAAADDIPNVLSQLAVVELYTDTSFWSVGARGALWE 200
 QY 179 CCATLGAEFOYQAQSKPKVEELNVICNVSOFSVKNPKGYKGVAFPLPTDAGVATATGKS 238
 DB 201 CGCATLGASFOYQAQSKPKVEELNVICNAAEFTINPKGVGQEFPLNKAGTVSATDTKD 260

QY 239 ATINHEWQVGSLSYRLNSLPYIGVQWRSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 298
 DB 261 ASIDYHEWQASLSYRLNMFTPIYIGVQWRSRATFDADNIRIAQPKLPTAVNLNTAWNPSL 320
 QY 299 LGNATALSTTDSFSDFMQIVSCQINKFKSRKACGTVGATLVADKWSLTAEARLINERA 358
 DB 321 SGSGIDVDT--KITDTLQIVSLQNLNKKSRKSCGLAIGTTIVDADKYAVTVETRLIDERA 378
 QY 359 AHVSGQFRF 367
 DB 379 AHVNAQFRF 387

Search completed: February 7, 2002, 21:38:11
 Job time: 7100 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:47 ; Search time 76.51 seconds
(without alignments)
175.872 Million cell updates/sec

Title: US-09-391-606-16

Perfect score: 1952

Sequence: 1 MLFVGNPSPSLIDTWE.....TAEARLINERAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1947	99.7	389	1	OMP1_CHLPN
2	1882	96.4	389	1	OM1N_CHLPN
3	1732	88.7	333	1	OM1K_CHLPN
4	1534.5	78.6	389	1	OM1A_CHLPS
5	1477	75.7	392	1	OM1P_CHLPS
6	1415	72.5	402	1	OM1E_CHLPS
7	1358.5	69.6	387	1	OMP1_CHLMU
8	1340.5	68.7	397	1	OM1H_CHLTR
9	1338.5	68.6	397	1	OM1N_CHLTR
10	1325	67.9	396	1	OM1A_CHLTR
11	1315.5	67.4	393	1	OM1E_CHLTR
12	1315.5	67.4	397	1	OM1C_CHLTR
13	1313.5	67.3	393	1	OM1L_CHLTR
14	1308.5	67.0	393	1	OM1D_CHLTR
15	1307.5	67.0	395	1	OM1F_CHLTR
16	1295	66.3	394	1	OM1M_CHLTR
17	1275	65.3	394	1	OM1B_CHLTR
18	1189	9.7	340	1	OMP2_CHLTR
19	182	9.3	340	1	OMP2_CHLMU
20	147.5	7.6	344	1	OMP2_CHLPN
21	100.5	5.1	1919	1	HEX_ADE12
22	98	5.0	1848	1	CBPA_CLOC
23	96	4.9	556	1	PHL_LEPIN
24	96	4.9	3421	1	TEGU_HSVB
25	95.5	4.9	864	1	AGLU_MUCJA
26	93.5	4.8	2021	1	OMPA_RICCN
27	93.5	4.8	2292	1	POLG_EMCVB
28	93.5	4.8	2292	1	POLG_EMCVD
29	92.5	4.7	1169	1	YR82_YEAST
30	92.5	4.7	2290	1	POLG_EMCV
31	91.5	4.7	639	1	AMYG_ASPPH
32	91	4.7	448	1	FADL_ECOLI
33	90.5	4.6	1086	1	NNTM_MOUSE

RESULT 1

ID	OMP1_CHLPN	STANDARD;	PRT;	389 AA.
AC	P27455; O9JOF6;			
DT	01-AUG-1992 (Rel. 23, Created)			
DT	01-AUG-1992 (Rel. 23, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).			
GN	OMPA OR OMP1 OR CPN0695 OR CP0051.			
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
OX	NCBI_TaxID=83558;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=IOL-207;			
RX	MEDLINE=91237311; PubMed=2033374;			
RA	Carter M.W., Al-Mahdawi S.A.H., Giles I.G., Trehan J.D.,			
RA	Ward M.E., Clarke I.N.;			
RT	"Nucleotide sequence and taxonomic value of the major outer membrane			
RT	protein gene of Chlamydia pneumoniae IOL-207.";			
RL	J. Gen. Microbiol. 137:465-475(1991).			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=TWAR;			
RX	MEDLINE=91244474; PubMed=1840574;			
RA	Perez Melgosa M., Kuo C.-C., Campbell L.A.;			
RT	"Sequence analysis of the major outer membrane protein gene of			
RT	Chlamydia pneumoniae.";			
RL	Infect. Immun. 59:2195-2199(1991).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RA	Mitchell W.M., Tharp A.C., Stratton C.W., Sriram S.;			
RL	Submitted (FEB-1999) to the EMBL/GenBank/DBJ databases.			
RN	[4]			
RP	SEQUENCE FROM N.A.			
RX	STRAIN=CWL029;			
RX	MEDLINE=99206606; PubMed=10192388;			
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,			
RA	Olinger L., Greenwood J., Davis R.W., Stephens R.S.;			
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";			
RL	Nat. Genet. 21:385-389(1999).			
RN	[5]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AR39;			
RX	MEDLINE=20150255; PubMed=10684935;			
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,			
RA	White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,			
RA	Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,			
RA	Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,			
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			
RT	pneumoniae AR39.";			
RL	Nucleic Acids Res. 28:1397-1406(2000).			
RN	[6]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=J138;			

34	90.5	4.6	1140	1	YHJL_ECOLI
35	90	4.6	533	1	FCY2_YEAST
36	89.5	4.6	725	1	FATA_VIBAN
37	89.5	4.6	867	1	SFMD_ECOLI
38	89.5	4.6	901	1	POLG_ENMG3
39	89	4.6	208	1	Y396_RICPR
40	89	4.6	363	1	YRAK_ECOLI
41	88.5	4.5	561	1	HLXB_PROMI
42	88.5	4.5	639	1	AMYG_ASAPK
43	87.5	4.5	482	1	PUR8_YEAST
44	87.5	4.5	1403	1	BIRF_MOUSE
45	87	4.5	664	1	PLBI_YEAST

ALIGNMENTS

P37650	escherichia
P17064	saccharomyc
P11461	vibrio angu
P77468	escherichia
P32540	mengo encep
Q92dd5	rickettsia
P43319	escherichia
P16465	proteus mir
P23176	aspergillus
Q05911	saccharomyc
Q9J1b6	mus musculu
P39105	saccharomyc


```
QY 182 ATLGAEFOYQAKPKVEELNVICNSQFSVKNPKGKGVAFPLPTDAGVATATGKTSATI 241
|||||
Db 204 ATLGAEFOYQAKPKVEELNVICNSQFSVKNPKGKGVAFPLPTDAGVATATGKTSATI 263
|||||
QY 242 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 301
|||||
Db 264 NYHEQVQASLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGN 323
|||||
QY 302 ATALSTDSFSDPMQIVSQINKEKSRKACGVTGATLVADKWSLTAEARLINERAAHV 361
|||||
Db 324 ATAVSSDQFSDPMQIVSQINKEKSRKACGVTGATLVADKWSLTAEARLINERAAHI 383
|||||
QY 362 SQQFRF 367
|||||
Db 384 SQQFRF 389
|||||

RESULT 3
OMIK_CHLPN STANDARD; PRT; 333 AA.
ID OM1K_CHLPN AC Q9XB4;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN (MOMP) (FRAGMENT).
GN OMPA OR OMP1.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KOALA TYPE I;
RX MEDLINE=93123168; PubMed=8419295;
RA Kallenboeck B., Kousoulas K.G., Storz J.;
RT "Structures of and allelic diversity and relationships among the major
RT outer membrane protein (ompA) genes of the four chlamydial species."
RL J. Bacteriol. 175:487-502(1993).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: M73038; AAD38210.1;
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin.
FT NON_TER 1 333
FT NON_TER 333 333
SQ SEQUENCE 333 AA; 35811 MW; 204604512C4C3B3F CRC64;

Query Match 88.7%; Score 1732; DB 1; Length 333;
Best Local Similarity 97.3%; Pred No. 3.2e-136;
Matches 324; Conservative 4; Mismatches 5; Indels 0; Gaps 0;

QY 23 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKPTGSAANYTTAVDRP 82
|||||
Db 1 AGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKFTSMGAKPTGSAANYTTAVDRP 60
|||||
QY 83 NPAYNKLHDAEFMTNAGFIALNWDREDFVCTLGASNGYIRGNSTAFNLVGLFGVKGT 142
|||||
```

```
Db 61 NPAYNKLHDAEFMTNAGFIALNWDREDFVCTLGASNGYIRGNSTAFNLVGLFGVKGT 120
|||||
QY 143 VNANELPNVSLNSGVVELYTDTSFWSVGARGALWECCATLGAEFOYQAKPKVEELNV 202
|||||
Db 121 VAANELPNVSLNSGVVELYTDTSFWSVGARGALWECCATLGAEFOYQAKPKVEELNV 180
|||||
QY 203 ICNVSQFSVKNPKGKGVAFPLPTDAGVATATGKTSATINHEQVQASLSYRLNSLPVY 262
|||||
Db 181 ICNVAQFSVKNPKGKGVAFPLPTDAGVATATGKTSATINHEQVQASLSYRLNSLPVY 240
|||||
QY 263 IGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFSDPMQIVSQI 322
|||||
Db 241 IGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFSDPMQIVSQI 300
|||||
QY 323 NKEKSRKACGVTGATLVADKWSLTAEARLIN 355
|||||
Db 301 NKEKSRKACGVTGATLVADKWSLTAEARLIN 333
|||||
```

RESULT 4

```
OM1A_CHLPS STANDARD; PRT; 389 AA.
ID OM1A_CHLPS AC P16567;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
GN OMPA OR OMP1.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OVINE ENZOOTIC ABORTION ISOLATE S26/3;
RX MEDLINE=90128177; PubMed=2612883;
RA Herring A.J., Tan T.W., Baxter S., Inglis N.F., Dunbar S.;
RT "Sequence analysis of the major outer membrane protein gene of an
RT ovine abortion strain of Chlamydia psittaci."
RL FEMS Microbiol. Lett. 53:153-158(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=BOVINE ABORTION ISOLATE BAI;
RX MEDLINE=96189695; PubMed=8605581;
RA Griffiths P.C., Plater J.M., Martin T.C., Hughes S.L.,
RA Hughes K.J., Hewinson R.G., Dawson M.;
RT "Epizootic bovine abortion in a dairy herd: characterization of a
RT Chlamydia psittaci isolate and antibody response."
RL Br. Vet. J. 151:683-693(1995).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: X51859; CAA36152.1;
DR EMBL; L39020; AACB02850.1;
DR PIR; S08770; MMCWP3.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
```

FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41883 MW; 741B5A23ACDBB47 CRC64;

Query Match 78.6%; Score 1534.5; DB 1; Length 389;
Best Local Similarity 75.7%; Pred. No. 8.8e-120;
Matches 278; Conservative 39; Mismatches 49; Indels 1; Gaps 1;

QY 2 LPVGNPSDLLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDVNTKI 82

QY 62 S-MGAKPTGSA--AANYTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRFVCTLGASN 120
DB 83 TGMGAYPTGTAAATYKTPDRNPAYNKLHDAEFTNAGFIALNIWDRFVCTLGASN 142

QY 121 GYIRGNSTAFNLVGLFGVKTGTVNANLNPVLSNGVVELYDTDSFSWSVGARGALWECG 180
DB 143 GYFKASSAFNLVGLFGVKTGTVNANLNPVLSNGVVELYDTDSFSWSVGARGALWECG 202

QY 181 CATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGTSAT 240
DB 203 CATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGTSAT 262

QY 241 INYHEWQVGLSYRLNVLVPIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLG 300
DB 263 IRYHEWQVGLSYRLNVLVPIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLG 322

QY 301 NATALSTDTSDSFDMQIVSQINKEFKSRKACGTVTGATLVADKWSLTAEARLINERAH 360
DB 323 EATLDTSNKFDADFLQIASIQINKMKSRKACGTVTGATLVADKWSLTAEARLINERAH 382

QY 361 VSGQFRF 367
DB 383 MNAQFRF 389

RESULT 5
OM1E_CHLPS STANDARD; PRT; 392 AA.
ID OM1P_CHLPS Q00087;
AC Q00087;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
GN OMPA OR OMP1.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94103736; PubMed=8277245;
RT Storey C., Lusher M., Yates P., Richmond S.;
RL "Evidence for Chlamydia pneumoniae of non-human origin.";
J. Gen. Microbiol. 139:2621-2626(1993).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
EMBL; X61096; CAA43409.1; .

DR PIR: A40371; A40371.
DR PIR: S16137; S16137.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 392 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 392 AA; 42069 MW; 8B3C5D90BBA26DB CRC64;

Query Match 75.7%; Score 1477; DB 1; Length 392;
Best Local Similarity 74.1%; Pred. No. 5.1e-115;
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;

QY 2 LPVGNPSDLLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSSLIDGTIWEAGAGDCPCATWCDATSLRAGFYGDYVDFRILKVDVNTKI 82

QY 62 S-MGAKPTGSA--AANYTTAVDRPNPAYNKLHDAEFTNAGFIALNIWDRFVCTLGASN 118
DB 83 SGMAAPTAASGTASNTTVAADRSNFAKGLQDAEFTNAGFIALNIWDRFVCTLGASN 142

QY 119 SNGYIRGNSTAFNLVGLFGVKTGTVNANLNPVLSNGVVELYDTDSFSWSVGARGALWE 178
DB 143 SNGYFKASSAFNLVGLFGVKTGTVNANLNPVLSNGVVELYDTDSFSWSVGARGALWE 201

QY 179 CGCATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGTF 236
DB 202 CGCATLGAEFQYAOQKPKVEELNVCNVSQFVNKPKYKGVAFPLPTDAGVATATGTF 261

QY 237 KSATINYNHEWQVGLSYRLNVLVPIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 296
DB 262 KSATINYNHEWQVGLSYRLNVLVPIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNP 321

QY 297 SLLGNATLSTDTSDSFDMQIVSQINKEFKSRKACGTVTGATLVADKWSLTAEARLINE 356
DB 322 TLLGVATTLDTSNKYADEFMQIVSQINKEFKSRKACGTVTGATLVADKWSLTAEARLINE 381

QY 357 RAAHVSQGF 367
DB 382 RAAHVSQGF 392

RESULT 6
OM1E_CHLPS STANDARD; PRT; 402 AA.
ID OM1E_CHLPS P10332;
AC P10332;
DT 01-MAR-1989 (Rel. 10, Created)
DT 01-MAR-1989 (Rel. 10, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
GN OMPA OR OMP1.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia phila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=EAE A22/M;
RA Pickett M.A., Everson S.J., Clarke I.N.;
RT "Chlamydia psittaci ewe abortion agent: complete nucleotide sequence of the major outer membrane protein gene.";
RL FEMS Microbiol. Lett. 55:229-234(1988).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; X12647; CAA31177.1; -
 EMBL; M36703; AAA23146.1; -
 PIR; S05954; MNCWPM.
 InterPro: IPR000604; Chlamydia_OMP.
 Pfam; PF01308; Chlamydia_OMP; 1.
 ProDom; PD001717; Chlamydia_OMP; 1.
 Outer membrane; Transmembrane; Porin; Signal.
 SIGNAL 1 22
 CHAIN 23 402 MAJOR OUTER MEMBRANE PROTEIN.
 E6CF00D9DF1E87A CRC64;
 SEQUENCE 402 AA; 43277 MW; 56CF00D9DF1E87A CRC64;

Query Match 72.5%; Score 1415; DB 1; Length 402;
 Best Local Similarity 70.3%; Pred. No. 7.2e-110;
 Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWEGACGDCPCATWCDAISLRAGFYGVDFRILKVDAPKTF 61
 DB 23 LPVGNPAEPSLLIDGTWEGACGDCPCATWCDAISLRAGFYGVDFRILKVDVKNKF 82
 QY 62 S-MGAKP---TGSAA-ANYTTAVDRNPAYNKLHDAEFTNAGFTALNLDWDFVFCFL 116
 DB 83 SGMAATPTQATGNASNTNPEANGRNIAIGHMQDAEWFSAFALNLDWDFDFCFL 142
 QY 117 GASNGYIRGNSTAFNLVGLFGVKGTTVNANE----LPNVSLNSGVVVELYDTFSFSVGA 172
 DB 143 GASNGYFKSSAAFNVLVGLIGESATSTSTELPMQLPNVGVITQGVVEFTDTFSFSVGA 202
 QY 173 RGALWECGATGAEEFYAQSKPKVEELNVCNVSFVSNKPKYKGA--PPLPDAGV 230
 DB 203 RGALWECGATGAEEFYAQSKPKVEELNVCNVSFVSNKPKYKGA--PPLPDAGV 262
 QY 231 ATATGPKSATINHEVQVQASISRLNSLVPYIGVQMSRATFDADNIRIAQKPLTAVLN 290
 DB 263 TEATDTPKSAITIKYHEVQVGLALSYRLNMLVPIGVNWSRATPDADTIRIAQKPKSEIN 322
 QY 291 LTAWNPISLGNATLSTT---DSFDFMQIVSQINKFKSRKACGVTVGATLVADKWSL 347
 DB 323 ITTWNPSLGSSTTTLPNNGKDLVSDVLAQIAQINKMKSRKACGVAVGATLVADKWSI 382
 QY 348 TEARLINERAHVSGQFRF 367
 DB 383 TGEARLINERAHVNAQFRF 402

RESULT 7
 OMP1_CHLMU STANDARD; PRT; 387 AA.
 AC P75024; Q04063; Q9X718;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP).
 GN OMPA OR OMP1 OR TC0052.
 OS Chlamydia muridarum.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83360;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN;
 RX MEDLINE-92039057; PubMed-1937036;
 RA Fielder T.J., Pal S., Peterson E.M., la Maza L.M.;
 RT "Sequence of the gene encoding the major outer membrane protein of the mouse pneumonitis biovar of Chlamydia trachomatis.";
 RL Gene 106:137-138(1991).
 RN [2]
 RP SEQUENCE FROM N.A.

RC STRAIN-MOPN;
 RX MEDLINE-94104488; PubMed-8277858;
 RA Zhang Y.X., Fox J.G., Ho Y., Zhang L., Stills H.F., Smith T.F.;
 RT "Comparison of the major outer membrane protein (MOMP) gene of mouse pneumonitis (MoPn) and hamster SFPP strains of Chlamydia trachomatis with other Chlamydia strains.";
 RL Mol. Biol. Evol. 10:1327-1342(1993).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-SP.BV.MOUSE / NIGG II;
 RA Carter M.W., Giles I., Everson J.S., Clarke I.N.;
 RT "Chlamydia trachomatis mouse biovar: major outer membrane protein gene.";
 RL (In) Mardh P.A., la Placa M., Ward M. (eds.);
 Proceedings of the European society for chlamydia research and the second international symposium of Uppsala university centre for std research, pp.38-38, University of Uppsala, Uppsala (1992).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-MOPN / NIGG;
 RX MEDLINE-20150255; PubMed-10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F., White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S., Linher K., Weidman J., Khouli H., Craven B., Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.";
 RL Nucleic Acids Res. 28:1397-1406(2000).
 RN [5]
 RP SEQUENCE OF 37-375 FROM N.A.
 RC STRAIN-MOPN;
 RX MEDLINE-93123168; PubMed-8419295;
 RA Kaltenboeck B., Kousoulas K.G., Storz J.;
 RT "Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species.";
 RL J. Bacteriol. 175:487-502(1993).
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
 CC -!- SUBCELLULAR LOCATION: CELL WALL SURFACE.
 CC -!- MISCELLANEOUS: MOMP IS RESPONSIBLE FOR THE STRUCTURAL INTEGRITY OF THE EXTRA-CELLULAR INFECTIOUS ELEMENTARY BODY & THE DEVELOPMENTAL CONVERSION TO THE PLASTIC AND FRAGILE INTRACELLULAR RETICULATE BODY.
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; M64171; AAA23144.1; -
 EMBL; U60196; AAB07068.1; -
 EMBL; X63409; CAA45006.1; -
 EMBL; AE002272; AAF38941.1; -
 EMBL; M73044; AAD29101.1; -
 TIGR; TC0052;
 InterPro: IPR000604; Chlamydia_OMP.
 Pfam; PF01308; Chlamydia_OMP; 1.
 ProDom; PD001717; Chlamydia_OMP; 1.
 Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
 SIGNAL 1 22 BY SIMILARITY.
 CHAIN 23 387 MAJOR OUTER MEMBRANE PROTEIN.
 CONFLICT 118 118 F -> Y (IN REF. 5).
 CONFLICT 123 123 Y -> F (IN REF. 5).
 CONFLICT 198 198 L -> P (IN REF. 1).
 CONFLICT 204 204 A -> P (IN REF. 1).

SQ SEQUENCE 387 AA; 42009 MW; 4FD6FDC23248E0A2 CRC64;

Query Match 69.6%; Score 1358.5; DB 1; Length 387;
Best Local Similarity 66.7%; Pred. No. 3.3e-105;
Matches 246; Conservative 50; Mismatches 66; Indels 7; Gaps 3;

QY 2 LPVGNPSPSLLIDGTWEGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSPMDIGILWEGFGDPCDCTTWCDAISLURLGYGDFVDFRILKVDVNNKF 82

QY 62 SMGAKPTGSAANYTTA---VDRPNPAYNKHLDHAEFTNAGFIALNIWDRFDFVCTLG 118
DB 83 ENGAAPTGD--ADLTATPTASRENPAYGKHMQDAEMFTNAYMALNIWDRFDFVCTLG 140

QY 119 SNGYIRGNSAFNLVGLFGVGTNNANELPNVSLNSGVWVLYTDTFSWSVGARGALWE 178
DB 141 TSGYLKGNASAFNLVGLFGVGTNNANELPNVSLNSGVWVLYTDTAFWSVGARGALWE 200

QY 179 CGCATLGAFFOYAKSKPKVEELNVCNVSQFVNPKYGVAPPLPTDAGVATATGKS 238
DB 201 CCATLGLASFOYAKSKPKVEELNVCNVAEFTINKPKYGVQEFFLNKAGTVSATDTKD 260

QY 239 ATINHEWQVGSLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAWNPSL 298
DB 261 ASIDYHEWQASLALSYRLNMFYIGVKWSRASFDADTIRIAQPKLETSLKMTTWNPTI 320

QY 299 LGNATALSTDSFSDPMQIVSCQINKFKSKACGVTVGATLVADADKWSLTAEARLINERA 358
DB 321 SSGSGIDVDT--KITDTLQIVSLQNLNMRKSRKSCGLAIGTTIVDADKYAVTVETRLIDERA 378

QY 359 AHVSGQFRF 367
DB 379 AHVNAQFRF 387

RESULT 8
OM1H.CHLTR STANDARD; PRT; 397 AA.

AC P13467;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H PRECURSOR (MOMP).
GN OMPA OR OMP1H.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN-SEROVAR H;
RX MEDLINE=90045958; PubMed=2813066;
RA Hamilton P.T., Malinowski D.P.;
RT "Nucleotide sequence of the major outer membrane protein gene from Chlamydia trachomatis serovar H.";
RL Nucleic Acids Res. 17:8366-8366(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-H/UW-4;
RA Dean D.A.;
RT "Sequence analysis of the major outer membrane protein gene (ompA) of Chlamydia trachomatis.";
RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).

EMBL; X16007; CAA34145.1; --
DR EMBL; AF304857; AAG41415.1; --
DR PIR; S06589; MNCWTH. Chlamydia_OMP.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR H.
SQ SEQUENCE 397 AA; 42946 MW; 478ACE3808BF37BA CRC64;

Query Match 68.7%; Score 1340.5; DB 1; Length 397;
Best Local Similarity 65.1%; Pred. No. 1e-103;
Matches 244; Conservative 51; Mismatches 71; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLLIDGTWEGAGDPCDPCATWCDALISLRAGFYGDYVDFRILKVDAPKTF 61
DB 23 LPVGNPAEPSPMDIGILWEGFGDPCDPCATWCDALISMRVYGGDFVDFRILKTDVNNKF 82

QY 62 SMGAKPTGSAANY---TTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFDFVCTLG 117
DB 83 QMGAAPTTNDAAADLQNDPKTNVARNPAYGKHMQDAEMFTNAYMALNIWDRFDFVCTLG 142

QY 118 ASNGYIRGNSAFNLVGLFGVGTNNANELPNVSLNSGVWVLYTDTFSWSVGARG 174
DB 143 ATTGYLKGNASAFNLVGLFGVGTNNANELPNVSLNSGVWVLYTDTTFAWSVGARGA 202

QY 175 ALWEGCATLGAFFOYAKSKPKVEELNVCNVSQFVNPKYGVAPPLPTDAGVATAT 234
DB 203 ALWEGCATLGAFFOYAKSKPKVEELNVCNVAEFTINKPKYGVQEFFLNKAGTVSATDTKD 262

QY 235 GTSKATINHEWQVGSLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294
DB 263 GTKDASIDYHEWQASLALSYRLNMFYIGVKWSRASFDADTIRIAQPKLAELDVTTL 322

QY 295 NPSLLGNAT--ALSTDSFSDPMQIVSCQINKFKSKACGVTVGATLVADADKWSLTAEAR 352
DB 323 NPTIAGKTGVASGSDNDLADTMQIVSLQNLNMRKSRKSCGLAIGTTIVDADKYAVTVETR 382

QY 353 LINERAAHVSGQFRF 367
DB 383 LIDERAAHVNAQFRF 397

RESULT 9
OM1H.CHLTR STANDARD; PRT; 397 AA.

AC P23114;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3 PRECURSOR (MOMP).
GN OMPA OR OMP1L3.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=404 / SEROVAR L3;
RX MEDLINE=91285429; PubMed=2060793;
RA Fielder T.J., Peterson E.M., de la Maza L.M.;
RT "Nucleotide sequence of DNA encoding the major outer membrane protein of Chlamydia trachomatis serovar L3.";
RL Gene 101:159-160(1991).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH

CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: X55700; CAA39226.1; -
 CC PIR: JE0413; JE0413.
 CC InterPro: IPR000604; Chlamydia_OMP.
 CC Pfam: PF01308; Chlamydia_OMP; 1.
 CC ProDom: PD001717; Chlamydia_OMP; 1.
 CC Outer membrane; Transmembrane; Porin; Signal.
 CC SIGNAL 1 22 BY SIMILARITY.
 CC CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L3.
 CC SEQUENCE 397 AA; 42885 MW; FIDDCF09535C2595 CRC64;

Query Match 68.6%; Score 1338.5; DB 1; Length 397;
 Best Local Similarity 65.1%; Pred. No. 1.5e-103;
 Matches 244; Conservative 49; Mismatches 73; Indels 9; Gaps 3;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDPAKTF 61
 DB 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISRMVGYGDFVDFVLKTDVNFKEF 82
 QY 62 SNGAKPTGSAANY-----TTAVDRNPAYNKLHDAEFTNAGFTALNIWDRFVFCITLG 117
 DB 83 QMGAEPTTSDAGLSNDFTTNVARNPAYGKHMQDAEMFTNAAAYMALNIWDRFVFCITLG 142
 QY 118 ASNGYIRGNSAFNLVGLFGVKGTTVNANE---LPNVSLNSGVVELYDTTFSFWSVGARG 174
 DB 143 ATTGYLKGNSASFNLVGLFGTKTSTNFNTAKLVPTALNQAVVELYDTTTFANVSVGARA 202
 QY 175 ALWEGCGATLGAFOYAKSPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATAT 234
 DB 203 ALWEGCGATLGAFOYAKSPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATAT 262
 QY 235 GTSKATINYHEWQVGSLSYRLNLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294
 DB 263 GTKDASIDYHEWQASLSYRLNLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 322
 QY 295 NPSLL--GNATLSTDSFSDFMQIVSCQINFKSRKAGCVTVGATLVADKWSLTAEARL 352
 DB 323 NPTIAGKGVSVAGSGENSELADTMQIVSLQNLNKKMRKSCGIAVGTTIYDADKYAVTVETR 382
 QY 353 LINERAAHVSGQFRF 367
 DB 383 LIDERAAHVNAQFRF 397

RESULT 10
 OMIA_CHLTR STANDARD; PRT; 396 AA.
 AC P23732;
 DT 01-NOV-1991 (Rel. 20, Created)
 DT 01-NOV-1991 (Rel. 20, Last sequence update)
 DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A PRECURSOR (MOMP).
 GN OMPA OR OMP1A.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=SAI/OT / SEROVAR A;

RX MEDLINE=91045088; PubMed=2235504;
 RA Hayes L.J., Clarke I.N.;
 RT "Nucleotide sequence of the major outer membrane protein gene of
 RT Chlamydia trachomatis strain A/SAI/OT.";
 RL Nucleic Acids Res. 18:6136-6136(1990).
 CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
 CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
 CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
 CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
 CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
 CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
 CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
 CC
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M58938; AAA23141.1; -
 CC EMBL: N33635; AAA92785.1; -
 CC PIR: S12799; S12799.
 CC InterPro: IPR000604; Chlamydia_OMP.
 CC Pfam: PF01308; Chlamydia_OMP; 1.
 CC ProDom: PD001717; Chlamydia_OMP; 1.
 CC Outer membrane; Transmembrane; Porin; Signal.
 CC SIGNAL 1 22
 CC CHAIN 23 396 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR A.
 CC SEQUENCE 396 AA; 42877 MW; 2F9D3B0CE2D08162 CRC64;

Query Match 67.9%; Score 1325; DB 1; Length 396;
 Best Local Similarity 64.7%; Pred. No. 2e-102;
 Matches 242; Conservative 47; Mismatches 77; Indels 8; Gaps 3;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDPAKTF 61
 DB 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISRMVGYGDFVDFVLKTDVNFKEF 82
 QY 62 SNGAKPTGSAANY-----TTAVDRNPAYNKLHDAEFTNAGFTALNIWDRFVFCITLG 117
 DB 83 QMGAEPTTSDAGLSNDFTTNVARNPAYGKHMQDAEMFTNAAAYMALNIWDRFVFCITLG 142
 QY 118 ASNGYIRGNSAFNLVGLFGVKGTTV---NANLNPVSLNSGVVELYDTTFSFWSVGARG 174
 DB 143 ATTGYLKGNSASFNLVGLFGTKTQSSGDTANIVTALNQAVVELYDTTTFANVSVGARA 202
 QY 175 ALWEGCGATLGAFOYAKSPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATAT 234
 DB 203 ALWEGCGATLGAFOYAKSPKVEELNVCNVSQFVSNKPKYKGVAFPLPTDAGVATAT 262
 QY 235 GTSKATINYHEWQVGSLSYRLNLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 294
 DB 263 GTKDASIDYHEWQASLSYRLNLPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTAW 322
 QY 295 NPSLLGNLSTTDS--FSDFMQIVSCQINFKSRKAGCVTVGATLVADKWSLTAEARL 353
 DB 323 NPTIAGKGVSVAGSGENSELADTMQIVSLQNLNKKMRKSCGIAVGTTIYDADKYAVTVETR 382
 QY 354 LINERAAHVSGQFRF 367
 DB 383 LIDERAAHVNAQFRF 396

RESULT 11
 OMIE_CHLTR STANDARD; PRT; 393 AA.
 ID OMIE_CHLTR
 AC P17451;
 DT 01-AUG-1990 (Rel. 15, Created)
 DT 01-AUG-1990 (Rel. 15, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)

DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR E PRECURSOR (MOMP).
GN OMPA OR OMPIC.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RC SEQUENCE FROM N.A.
RP STRAIN=BOUR / SEROVAR E;
RX MEDLINE=90287737; PubMed=2356137;
RA Peterson E.M., Markoff B.A., de la Maza L.M.;
RT "The major outer membrane protein nucleotide sequence of Chlamydia trachomatis, serovar E.";
RL Nucleic Acids Res. 18:3414-3414(1990).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING. PERMITTING DIFFUSION OF SOLUTIONS THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X52557; CAA36791.1; -
DR PIR; S10201; MWCWTC.
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR E.
SQ SEQUENCE 393 AA; 42424 MW; AB2B82D16027B361 CRC64;

Query Match 67.4%; Score 1315.5; DB 1; Length 393;
Best Local Similarity 65.3%; Pred. No. 1.2e-101;
Matches 243; Conservative 46; Mismatches 76; Indels 7; Gaps 4;

QY 2 LPVGNPSPSLIDGTINWEGAGDPCDPCATWCDALSLRAGFYGDYVDFDRILKVDPAKTF 61
DB 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISLRMGYGGDFVDFRVLKTDVNEKF 82

QY 62 SMGAKP---TGSAAAYTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFDFVCTLGA 118
DB 83 QMGDKPTSTTGNATAP-TTLTARENPAYGRHMQDAEMFTNACMALNIWDRFDFVCTLGA 141

QY 119 SNGYIRGNSTARNVLGLFG--VKGTIVANLPLPNVSLNGVVELYTDTSFWSVSGARGAL 176
DB 142 SSGYLKGNASFNVLGLFGDNEQSTVTNSVNNSLDQSVVELYTDTSFWSVSGARGAL 201

QY 177 WECGCATLCAEFOYQAQSKPKVEELNVCNVSQFNKPKYKGVAFPLPTDAGVATATGT 236
DB 202 WECGCATLCAEFOYQAQSKPKVEELNVCNVAEFTNKPKYGVQEPFLALAGTDAATGT 261

QY 237 KSATINHEWQVAGSLYSRLNSLVPYIGVQWRSATFDADNIRIAQKPLTAVLNITANNP 296
DB 262 KDASIDYHEWQVAGSLYSRLNFTPIYGVKWSRASFDADTIRIAQKPSATAIFDTTLNP 321

QY 297 SLLGNATLSTTD-SFSDPMQIVSQINKFKSKAGGVTVGATLVADKWSITAEARLIN 355
DB 322 TAGAGDVKASAEQGLGDTMQIVSLQNLNKKSRKSCGTAAGTTIVDADKIAVTVETRLID 381

QY 356 ERAAHVSGQFRF 367
DB 382 ERAAHVNAQFRF 393

RESULT 12
OMIC_CHLTR ID OMC_CHLTR STANDARD; PRT; 397 AA.
AC P08780;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR C PRECURSOR (MOMP).
GN OMPA OR OMPIC.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN [1]
RC SEQUENCE FROM N.A.
RP MEDLINE=87307955; PubMed=3040664;
RX Stephens R.S., Sanchez-Pescador R., Wagar E.A., Inouye C., Urdea M.S.;
RA "Diversity of Chlamydia trachomatis major outer membrane protein genes.";
RT J. Infect. Dis. 169:3879-3885(1987).
RL J. Bacteriol. 169:3879-3885(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C/TW3;
RX MEDLINE=20407420; PubMed=10950788;
RA Dean D., Suchland R.J., Stamm W.E.;
RT "Evidence for long-term cervical persistence of Chlamydia trachomatis by omp1 genotyping.";
RL J. Infect. Dis. 182:909-916(2000).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY BODIES AND PORIN FORMING. PERMITTING DIFFUSION OF SOLUTIONS THROUGH THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M17343; AAA23156.1; -
DR EMBL; AF202455; AAG09443.1; -
DR PIR; S11011; MWCWTC.
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD001717; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 397 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR C.
SQ SEQUENCE 397 AA; 42892 MW; 0047BCDB108E5309 CRC64;

Query Match 67.4%; Score 1315.5; DB 1; Length 397;
Best Local Similarity 64.3%; Pred. No. 1.2e-101;
Matches 241; Conservative 49; Mismatches 76; Indels 9; Gaps 3;

QY 2 LPVGNPSPSLIDGTINWEGAGDPCDPCATWCDALSLRAGFYGDYVDFDRILKVDPAKTF 61
DB 23 LPVGNPAEPLSMIDGILWEGFGDPCDCTTWCDAISLRMGYGGDFVDFRVLKTDVNEKF 82

QY 62 SMGAKPTGSAANY----TTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFDFVCTIG 117
DB 83 QMGAAPTTSDVAGLQNDPTINVARPNPAYKHMQDAEMFTNAAAYMALNIWDRFDFVCTIG 142

QY 118 ASGYIRGNSTARNVLGLFGVKGTTVNANE----LPNVSLNGVVELYTDTSFWSVSGARG 174
DB 143 ATTGYLKNASFNVLGLFGTKTQSSSFNTAKLIPNTALNEAVVELYINTTFAWSVGARA 202

QY 175 ALWECGCATLGAEFOYQAQSKPKVEELNVCNVSQFNKPKYKGVAFPLPTDAGVATAT 234

```
Db 203 ALWECGCATLGASFOYASQSKPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTEAAT 262
Qy 235 GTSKATINHEWQVGSLSYRLNSLVPIGVQWSRATDADNIRIAQPKLPTAVLNLTAW 294
Db 263 GTKDASIDYHEWQASLALSRYLNMFTPIGVKWSRVSFDADTIRIAQPKLAEALIDVTL 322
Qy 295 NPSLLGNATALS--TTDSFDMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEAR 352
Db 323 NRTTAGKSVVSAGTNDLADMTQVLSQLNKMRSKSCGIAVGTITVDADKYAVTVEAR 382
Qy 353 LINERAAHVSGQFR 367
Db 383 LIDERRAAHVNAQFR 397

RESULT 13
OM1D_CHLTR STANDARD; PRT; 393 AA.
AC P19542;
DT 01-FEB-1991 (Rel. 17, Created)
DT 01-FEB-1991 (Rel. 17, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1 PRECURSOR (MOMP).
GN OMPA OR OMP1L1.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN NCBI_TaxID=813;
RP SEQUENCE FROM N.A.
RA Pickett M.A., Ward M.E., Clarke I.N.;
RT "Complete nucleotide sequence of the major outer membrane protein
  gene from Chlamydia trachomatis serovar L1.";
RL FEMS Microbiol. Lett. 42:185-190(1987).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
  BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
  THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
  MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M36533; AAA23142.1; -
DR PIR; S06259; S06259.
DR InterPro; IPR000604; Chlamydia_OMP.
DR Pfam; PF01308; Chlamydia_OMP; 1.
DR ProDom; PD00117; Chlamydia_OMP; 1.
KW Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 22
FT CHAIN 23 393 MAJOR OUTER MEMBRANE PROTEIN, SEROVAR L1.
SQ SEQUENCE 393 AA; 42543 MW; 7A952839408EE2DF CRC64;

Query Match 67.3%; Score 1313.5; DB 1; Length 393;
Best Local Similarity 64.5%; Pred. No. 1.8e-101;
Matches 240; Conservative 51; Mismatches 74; Indels 7; Gaps 4;

Qy 2 LPVGNPSPSLIDGTIWEAGAGDPCDCAWCAISLRAGFYGVDFRILKVDAPKTF 61
Db 23 LPVGNPAEPRLMIDGILWEGFGDPCDCTTWCDAISLRMGYGVDFRVLQTDVNEK 82

Qy 62 SNGAKP---TCSAANYTAVDRPNPAYNKHLDAEWTNAGFTALNTWDRFDVCTLGA 118
Db 83 QMGAKPTATTGNAAP-STCTARENPAYGRHMQDAEMFTNAYMALNTWDRFDVCTLGA 141

Qy 119 SNGYIRGNSTAFNLVGLFG--VKGTTVNANELPNVSLNSGVVELYTDTSFMSVSGARGAL 176
```

```
Db 142 TSGYLKNGSASFNLVGLFGDNEQSTYKKNDAVPNMSFQSVVELYTDITTFWSVGAARAL 201
Qy 177 WECCCATLGAFQYASQPKPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTATGT 236
Db 202 WECCCATLGASFOYASQSKPKVEELNVLNASEFTINKPKYGVGAEPFLNITAGTATGT 261
Qy 237 KSATINHEWQVGSLSYRLNSLVPIGVQWSRATDADNIRIAQPKLPTAVLNLTAWNP 296
Db 262 KDSIDYHEWQASLALSRYLNMFTPIGVKWSRVSFDADTIRIAQPKLAEALIDVTLNLP 321
Qy 297 SLLGNATALSTD--SFDFMQIVSCQINKFKSRKACGVTVGATLVADKWSLTAEARLIN 355
Db 322 TIAGAGEVKAEGQLGDTMQVLSQLNKMRSKSCGIAVGTITVDADKYAVTVETRLID 381
Qy 356 EFAAHVSGQFR 367
Db 382 EFAAHVNAQFR 393

RESULT 14
OM1D_CHLTR STANDARD; PRT; 393 AA.
AC Q46409;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN, SEROVAR D PRECURSOR (MOMP).
GN OMPA OR OMP1 OR CT681.
OS Chlamydia trachomatis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=813;
RN NCBI_TaxID=813;
RP SEQUENCE FROM N.A.
RA STRAIN=D/B-120;
RT MEDLINE=93013014; PubMed=1398119;
RA Sayada C., Denamur E., Elion J.;
RT "Complete sequence of the major outer membrane protein-encoding gene
  of Chlamydia trachomatis serovar Da.";
RL Gene 120:129-130(1992).
CC [2]
CC SEQUENCE FROM N.A.
CC STRAIN=D/IU-71960;
CC MEDLINE=98339860; PubMed=9673241;
CC Stothard D.R., Boguslawski G., Jones R.B.;
CC "Phylogenetic analysis of the Chlamydia trachomatis major outer
  membrane protein and examination of potential pathogenic
  determinants.";
CC Infect. Immun. 66:3618-3625(1998).
CC [3]
CC SEQUENCE FROM N.A.
CC STRAIN=D/UW-3/CX;
CC MEDLINE=99000809; PubMed=9784136;
CC Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
CC Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
CC Davis R.W.;
CC "Genome sequence of an obligate intracellular pathogen of humans:
  Chlamydia trachomatis.";
CC Science 282:754-759(1998).
CC -!- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
  BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
  THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC -!- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
  MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
CC -!- SIMILARITY: BELONGS TO THE CHLAMYDIAL OMP FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation
  at the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@isb-sib.ch).
CC -----
```

or send an email to license@isb-sib.ch).

or send an email to license@isb-sib.ch).

CC	EMBL; X62918; CAA44701.1; -
CC	EMBL; AF063195; AAC31436.2; -
DR	EMBL; AE001338; AAC68276.1; -
DR	InterPro: IPR000604; Chlamydia_OMP.
DR	Pfam: PF01308; Chlamydia_OMP; 1.
DR	ProDom: PD001717; Chlamydia_OMP; 1.
KW	Outer membrane; Transmembrane; Porin; Signal; Complete proteome.
FT	SIGNAL 1 22 BY SIMILARITY.
FT	CHAIN 23 393
SQ	SEQUENCE 393 AA; 42438 MW; 8CD692FD3EFF21D6 CRC64;

-- SIMILARITY: BELONGS TO THE CHLAWYDIAL OMP FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL oustation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>) or send an email to license@sib-sib.ch.

Search completed: February 7, 2002, 21:42:47
Job time: 601 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:21 ; Search time 172 Seconds
(without alignments)
312.104 Million cell updates/sec

Title: US-09-391-606-16

Perfect score: 1952

Sequence: 1 MLPVGNPSPDLLIDGTWE.....TAEARLINERAHVSGQFRF 367

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	1910	97.8	389	2 Q08085	Q08085 chlamydia p
2	1531.5	78.5	389	2 Q9APM4	Q9APM4 chlamydogphi
3	1529.5	78.4	388	2 Q9AIK1	Q9AIK1 chlamydia p
4	1501.5	76.9	391	2 Q46235	Q46235 chlamydia p
5	1497.5	76.7	388	2 Q9AIK0	Q9AIK0 chlamydia p
6	1497.5	76.7	389	2 Q9AIH9	Q9AIH9 chlamydogphi
7	1487	76.2	392	2 Q9AIJ4	Q9AIJ4 chlamydia p
8	1480	75.8	381	2 Q9AIJ2	Q9AIJ2 chlamydogphi
9	1479	75.8	390	2 Q9AIJ5	Q9AIJ5 chlamydia p
10	1477	75.7	392	2 Q99QB0	Q99QB0 chlamydogphi
11	1469.5	75.3	380	2 Q9AIJ1	Q9AIJ1 chlamydogphi
12	1423	72.9	391	2 Q9AIJ2	Q9AIJ2 chlamydia p
13	1420.5	72.8	341	2 Q9X717	Q9X717 chlamydogphi
14	1419	72.7	402	2 Q46193	Q46193 chlamydia p
15	1415	72.5	382	2 Q9AIJ9	Q9AIJ9 chlamydia p
16	1415	72.5	395	2 Q9AIJ7	Q9AIJ7 chlamydia p
17	1415	72.5	397	2 Q9AIJ8	Q9AIJ8 chlamydia p
18	1413	72.4	402	2 Q9AIJ6	Q9AIJ6 chlamydia p
19	1410	72.2	402	2 Q9AIJ0	Q9AIJ0 chlamydia p

20	1408	72.1	402	2 Q46203	Q46203 chlamydia p
21	1407	72.1	402	2 Q46236	Q46236 chlamydia p
22	1382	70.8	402	2 Q9AIJ3	Q9AIJ3 chlamydia p
23	1371.5	70.3	352	2 Q70085	Q70085 chlamydia p
24	1367.5	70.1	352	2 Q70050	Q70050 chlamydia p
25	1363.5	69.9	352	2 Q69307	Q69307 chlamydia p
26	1362.5	69.8	352	2 Q69306	Q69306 chlamydia p
27	1357	69.5	337	2 Q9XBF6	Q9XBF6 chlamydogphi
28	1353	69.3	336	2 Q9XBF5	Q9XBF5 chlamydogphi
29	1353	69.3	353	2 Q69305	Q69305 chlamydia p
30	1352	69.3	353	2 Q9AIJ8	Q9AIJ8 chlamydia s
31	1351	69.2	404	2 Q46407	Q46407 chlamydia t
32	1343	68.8	389	2 Q9AIJ4	Q9AIJ4 chlamydia s
33	1342	68.8	386	2 Q69093	Q69093 chlamydia t
34	1341.5	68.7	386	2 Q9AIJ5	Q9AIJ5 chlamydia s
35	1339.5	68.6	397	2 Q69094	Q69094 chlamydia t
36	1337.5	68.5	397	2 Q69095	Q69095 chlamydia t
37	1336.5	68.5	387	2 Q9AIJ0	Q9AIJ0 chlamydia s
38	1335.5	68.4	397	2 Q9F951	Q9F951 chlamydia t
39	1333	68.3	340	2 Q9XBF2	Q9XBF2 chlamydogphi
40	1331.5	68.2	387	2 Q9AIJ1	Q9AIJ1 chlamydia s
41	1331.5	68.2	397	2 Q9F950	Q9F950 chlamydia t
42	1330.5	68.2	385	2 Q9AIJ6	Q9AIJ6 chlamydia s
43	1328	68.0	396	2 Q46406	Q46406 chlamydia t
44	1327.5	68.0	385	2 Q9AIJ7	Q9AIJ7 chlamydia s
45	1326	67.9	356	2 Q52924	Q52924 chlamydia p

ALIGNMENTS

RESULT 1

Q08085 ID Q08085 PRELIMINARY; PRT; 389 AA.
AC Q08085;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (MOMP)
OS Chlamydia psittaci (Chlamydogphila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydogphila.
OX NCBI_taxID=83554;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=KOALA TYPE 1;
RX MEDLINE=94171025; PubMed=8125292;
RA Girjes A.A., Carrick F.N., Lavin M.F.;
RT Remarkable sequence relatedness in the DNA encoding the major outer
RT membrane protein of Chlamydia psittaci (koala type I) and Chlamydia
RT pneumoniae.";
RL Gene 138:139-142(1994).
CC !- FUNCTION: STRUCTURAL RIGIDITY OF THE OUTER MEMBRANE OF ELEMENTARY
CC BODIES AND PORIN FORMING, PERMITTING DIFFUSION OF SOLUTES THROUGH
CC THE INTRACELLULAR RETICULATE BODY MEMBRANE.
CC !- SUBUNIT: DISULFIDE BOND INTERACTIONS WITHIN AND BETWEEN MOMP
CC MOLECULES & OTHER COMPONENTS FORM HIGH MOLECULAR-WEIGHT OLIGOMERS.
CC !- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. OUTER MEMBRANE.
DR EMBL; X72023; CAA50906.1; -
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam: PF01308; Chlamydia_OMP; 1.
DR PRINTS; PR01334; CHLAMIDI_OMP.
DR PRODOM; PD001717; Chlamydia_OMP; 1.
DR Outer membrane; Transmembrane; Porin; Signal.
FT SIGNAL 1 23 BY SIMILARITY.
FT CHAIN 24 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41579 MW; 5DC50E85A6F4E50F CRC64;

Query Match 97.8%; Score 1910; DB 2; Length 389;
Best Local Similarity 97.5%; Pred. No. 4.4e-147;
Matches 357; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

Oy 2 LPVGNPSPDLLIDGTWEAGAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61

```
|||||
Db 24 LPVGNPDPSSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 83
Qy 62 SMGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCITLGSNG 121
Db 84 SMGAKPTGSATANYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCITLGSNG 143
Qy 122 YIRGNSTAFNLVGLFGYKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 181
Db 144 YIKGNSTAFNLVGLFGYKGTVAANELPNVSLNGVVELYTDTSFWSVGARGALWECG 203
Qy 182 ATLGAEFQYQASPKPVEELNVCNVSQVSNKPKGYKGVAFPLPTDAGVATATGKTSAT I 241
Db 204 ATLGAEFQYQASPKPVEELNVCNVAQSFVNKPKGYKGVAFPLPTDAGVATATGKTSAT I 263
Qy 242 NYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNLTAWNPSLLGN 301
Db 264 NYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNLTAWNPSLLGN 323
Qy 302 ATALSTTDSFDMQIVSCQINKFKSRKACGVTVCATLVADADKWSLTAEARLINERAHV 361
Db 324 TTTLATSDSFDMQIVSCQINKFKSRKACGVTVCATLVADADKWSLTAEARLINERAHV 383
Qy 362 SQGFRR 367
Db 384 SQGFRR 389

RESULT 2
Q9APM4 PRELIMINARY; PRT; 389 AA.
ID Q9APM4
AC Q9APM4; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMPL.
OS Chlamydia abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20569239; PubMed=11119563;
RA Vreton E., Psarrou E., Kaisar M., Vlisidou I., Salti-Montesanto V.,
RT Longbottom D.;
RT Identification of protective epitopes by sequencing of the major
RT outer membrane protein gene of a variant strain of Chlamydia psittaci
RL Infect. Immun. 69:607-612(2001).
DR EMBL: AF272945; AAG53881.1; -
KW Signal.
FT SIGNAL.
FT CHAIN 1 22 POTENTIAL.
FT CHAIN 23 389 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 389 AA; 41897 MW; 2051369C7DBAA5 CRC64;

Query Match 78.5%; Score 1531.5; DB 2; Length 389;
Best Local Similarity 75.7%; Pred. No. 2.3e-116;
Matches 278; Conservative 38; Mismatches 50; Indels 1; Gaps 1;

Qy 2 LPVGNPDPSSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
Db 23 LPVGNPAEPSSLIDGTWEGASGDDPCDPCSTWCDAISLRAGYGYGVDFRILKVDVNKTI 82
Qy 62 S-MGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCITLGSN 120
Db 83 TGMGAVPTGTAADYKTPDRENPAYNKHLDHAEFTNAGFTALNIWDRFVFCITLGSN 142
Qy 121 GYIRGNSTAFNLVGLFGYKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 180
Db 143 GYFKASSAFAFLVGLFGYKGTSSVAADQLPNVGITQIGVEFYDTDTFWSVGARGALWECG 202
```

```
Qy 181 CATLGAEFQYQASPKPVEELNVCNVSQVSNKPKGYKGVAFPLPTDAGVATATGKTSAT 240
Db 203 CATLGAEFQYQASPKPVEELNVCNVSQVSNKPKGYKGVAFPLPTDAGVATATGKTSAT 262
Qy 241 INYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNLTAWNPSLLG 300
Db 263 IKYHEWQVGLSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLAAAVALNLTWNPTLLG 322
Qy 301 NATALSTTDSFDMQIVSCQINKFKSRKACGVTVCATLVADADKWSLTAEARLINERAHV 360
Db 323 EATLTDNNKFAFQLQIASIQINKMKSRKACGAVGATLIDADKWSITGEARLINERAHV 382
Qy 361 VSGOFRF 367
Db 383 MNAOFRF 389

RESULT 3
Q9AIK1 PRELIMINARY; PRT; 388 AA.
ID Q9AIK1
AC Q9AIK1; 2001 (Tremblrel. 17, Created)
DT 01-JUN-2001 (Tremblrel. 17, Last sequence update)
DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=VS225;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL: AF269259; AAK00240.1; -
KW Signal.
FT NON_TER 1 1 POTENTIAL.
FT SIGNAL <1 19
FT CHAIN 20 388 MAJOR OUTER MEMBRANE PROTEIN.
SQ SEQUENCE 388 AA; 41573 MW; 8E232D2C9B9948D CRC64;

Query Match 78.4%; Score 1529.5; DB 2; Length 388;
Best Local Similarity 76.2%; Pred. No. 3.3e-116;
Matches 281; Conservative 35; Mismatches 50; Indels 3; Gaps 2;

Qy 2 LPVGNPDPSSLIDGTIWEGAAGDPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF 61
Db 20 LPVGNPAEPSSLIDGTWEGASGDDPCDPCATWCDAISLRAGYGYGVDFRILKVDVNKTI 79
Qy 62 S-MGAKPTGSAANYYTTAVDRNPAYNKHLDHAEFTNAGFTALNIWDRFVFCITLGSN 120
Db 80 SGMGAAPTGSAAADYKTPDRENPAYNKHLDHAEFTNAGFTALNIWDRFVFCITLGSN 139
Qy 121 GYIRGNSTAFNLVGLFGYKGTTVNANELPNVSLNGVVELYTDTSFWSVGARGALWECG 180
Db 140 GYFKASSAFAFLVGLFGYKGTSSVAADQLPNVGITQIGVEFYDTDTFWSVGARGALWECG 199
Qy 181 CATLGAEFQYQASPKPVEELNVCNVSQVSNKPKGYKGVAFPLPTDAGVATATGKTSAT 238
Db 200 CATLGAEFQYQASPKPVEELNVCNVSQVSNKPKGYKGVAFPLPTDAGVATATGKTSAT 259
Qy 239 ATINYHEWQVGSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLPTAVNLNLTAWNPSL 298
Db 260 ATLYKHEWQVGLSLSYRLNSLPVYIGVQWSRATFDADNIRIAQPKLAAAVALNLTWNPTL 319
Qy 299 LCNATALSTTDSFDMQIVSCQINKFKSRKACGVTVCATLVADADKWSLTAEARLINERA 358
Db 320 LGAEATLADSNKFCDFLQIASIQINKMKSRKACGAVGATLIDADKWSITGEARLINERA 379
Qy 359 AHVSGOFRF 367
```


DR		EMBL; AF269267; AAK00248.1; -;	
KW	SIGNAL.		
FT	SIGNAL	1 22	POTENTIAL.
CH	CHAIN	23 392	MAJOR OUTER MEMBRANE PROTEIN.
SQ	SEQUENCE	392 AA; 42293 MW; FC31FCO51955246C CRC64;	

		Query Match	76.2%; Score 1487; DB 2; Length 392;
		Best local similarity	74.2%; Pred.No.9.4e-113;
		Matches 276; Conservative 39; Mismatches 49; Indels 8; Gaps	

QY	2	LPVGNPSDPSLLIDGTIWEGAGGDCPCATWCDAISLRAGEYGDYVFDRILKVDAPKTF 61	: : : : : : : : : : : : : : : :
DB	23	LPVGNPAEPSLLIDGTIWEGASGGDCPCATWCDAISIRAGIYGDYFDRVLKVDVNKTF 82	: : : : : : : : : : : : : : ~
QY	62	SMGAK----PTGSAAANYTTAVDRPNPAYNKKHLHDAWEFTNAGFIALNIWDREFVICTLG 117	: : : : : : : :
DB	83	SGMKSPEATGTSAT-TTAVDTNLAYCKKLQDAWEFTNAFLALNIWDREDFICTLG 141	: : : : : : : :
QY	118	ASNGYTRNSTARNLVCLGFVKCTTVANNELPVSLSUGNVVELYTDSFSWSVGANGALW 177	: : : : : : :
DB	142	ASNGYFKASSAANLVLGLIKGTKDEN-NQLPNVAITQGVEFYTDTSFSWSVGANGALW 200	: : : : : : :
QY	178	ECCCATILGAEFYAQSAPPKEEVLNVICNVYSQFSVPKPQKGVA--FPLPTDAGVATATG 235	: : : : : : :
DB	201	ECCCATILGAEFYAQSAPNKIEMLNVSTSSPAQFVIHKPRGYKGTGSNFPIPDAGTHAAD 260	: : : : : : :
QY	236	TKSATINYHEWQVAGSLSYRLNSLYPYIGVOWSRATFDADNRIRIAOKPLTAVINLTAWN 295	: : : : : : :
DB	261	TKSATLKHYEWQVGLALSRYLNLMVYPYIGNWSRAIFDADTRIAPOKLATAVLDITTWN 320	: : : : : : :
QY	296	PSLLGNATALSTTDSFSDFMQIVSQCNPKFSRKACGVTVGATFLVDADKWSLTAEARLIN 355	: : : : : : :
DB	321	PTTLGKATTVDGNTYSDFIQLASIQINKMSRKACGVAVGATLIADKWSITGEARLIN 380	: : : : : : :
QY	356	ERAHVSGQRPF 367	: : : : : : :
DB	381	ERAHMNAQRF 392	: : : : : : :

RESULT 8

Q9AIL2	PRELIMINARY;	PRT; 381 AA.
ID AC Q9AIL2		
CD DT 01-JUN-2001	(TrEMBLrel. 17, Created)	
DD DT 01-JUN-2001	(TrEMBLrel. 17, Last sequence update)	
DT DT 01-JUN-2001	(TrEMBLrel. 17, Last annotation update)	
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).		
DN OMFA.		
OS Chlamydomophila pecorum.		
OX Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.		
CX NCBI_TaxID=85991;		
[1]		
RN RP SEQUENCE OF 34-369 FROM N.A.		
RC STRAIN=1710S;		
RC MEDLINE=93123168; PubMed=8419295;		
RA Kaltenboeck B., Kousoulas K.G.; Storz J.;		
RT "Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species.";		
RL Int. J. Syst. Evol. Microbiol. 51:203-220(1993).		
[2]		
RN RP SEQUENCE FROM N.A.		
RC STRAIN=1710S;		
RX MEDLINE=21078680; PubMed=11211261;		
RA Bush R.M.; Everett K.D.;		
RT "Molecular evolution of the Chlamydiaceae."		
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).		
[3]		
RN RP SEQUENCE FROM N.A.		
RC STRAIN=1710S;		
CA Everett K.D.E.; Hamblly W.A.; Andersen A.A.;		
RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.		
DR EMBL; AF269279; AAK00260.1; -;		


```

KW signal.
FT NON_TER 1 15
FT SIGNAL <1 15
FT CHAIN 16 381
SQ SEQUENCE 381 AA; 41332 MW; 29406725CF9D3512 CRC64;

Query Match
Best Local Similarity 75.8%; Score 1480; DB 2; Length 381;
Matches 267; Conservative 47; Mismatches 51; Indels 2; Gaps 2;

Qy 2 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVFDRLKVDAPKTF 61
Db 16 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVFDRLKVDAPKTF 75

Qy 62 SMGAKPTG-SAAANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNINWDRFDFVCTLG 120
Db 76 SMGAKPTG-SAAANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNINWDRFDFVCTLG 135

Qy 121 GYIRGNSTAFNLVGLFGVKGTTVNAVNLPNVSLNGVVELYDTDSFSVSGARGALWECG 180
Db 136 GYIRGNSTAFNLVGLFGVKGTTVNAVNLPNVSLNGVVELYDTDSFSVSGARGALWECG 194

Qy 191 CATLGAEFYQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTSKAT 240
Db 195 CATLGAEFYQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTSKAT 254

Qy 241 INYHEWQVGSLSYRLNSLVPYIGVOWSRATFDADNIRIAQPKLPTAVLNLTANNPSLLG 300
Db 255 INYHEWQVGSLSYRLNSLVPYIGVOWSRATFDADNIRIAQPKLPTAVLNLTANNPSLLG 314

Qy 301 NATALSTTDSFDMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAEARLNERAAH 360
Db 315 NATALSTTDSFDMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAEARLNERAAH 374

Qy 361 VSGQFRF 367
Db 375 VSGQFRF 381

RESULT 9
Qy 09AIJ5 PRELIMINARY; PRT; 390 AA.
AC 09AIJ5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydia psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NEW JERSEY 1, N11;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
DR Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269266; AAK0247.1; -.
DR EMBL; AF269266; AAK0247.1; -.
KW Signal.
FT SIGNAL 1 20
FT CHAIN 21 390
SQ SEQUENCE 390 AA; 42042 MW; B62858403BFA456 CRC64;

Query Match
Best Local Similarity 75.8%; Score 1479; DB 2; Length 390;
Matches 274; Conservative 41; Mismatches 49; Indels 8; Gaps 4;

Qy 2 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVFDRLKVDAPKTF 61
Db 16 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVFDRLKVDAPKTF 75

Qy 62 SMGAKPTG-SAAANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNINWDRFDFVCTLG 120
Db 76 SMGAKPTG-SAAANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNINWDRFDFVCTLG 135

Qy 121 GYIRGNSTAFNLVGLFGVKGTTVNAVNLPNVSLNGVVELYDTDSFSVSGARGALWECG 180
Db 136 GYIRGNSTAFNLVGLFGVKGTTVNAVNLPNVSLNGVVELYDTDSFSVSGARGALWECG 194

Qy 191 CATLGAEFYQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTSKAT 240
Db 195 CATLGAEFYQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTSKAT 254

Qy 241 INYHEWQVGSLSYRLNSLVPYIGVOWSRATFDADNIRIAQPKLPTAVLNLTANNPSLLG 300
Db 255 INYHEWQVGSLSYRLNSLVPYIGVOWSRATFDADNIRIAQPKLPTAVLNLTANNPSLLG 314

Qy 301 NATALSTTDSFDMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAEARLNERAAH 360
Db 315 NATALSTTDSFDMQIVSCQINKFKSRKACGVTVGATLVLDADKWSLTAEARLNERAAH 374

Qy 361 VSGQFRF 367
Db 375 VSGQFRF 381

RESULT 10
Qy 09Q9BO PRELIMINARY; PRT; 392 AA.
AC 09Q9BO;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMPA.
OS Chlamydia felis.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83556;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-FP BAKER, ATCC VR120, AND FP CELLO;
RX MEDLINE=21078680; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
DR Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL; AF269257; AAK00238.1; -.
DR EMBL; AF269258; AAK00239.1; -.
KW Signal.
FT SIGNAL 1 22
FT CHAIN 23 392
SQ SEQUENCE 392 AA; 42051 MW; 88B3C09C1FEE26DB CRC64;

Query Match
Best Local Similarity 75.7%; Score 1477; DB 2; Length 392;
Matches 275; Conservative 38; Mismatches 52; Indels 6; Gaps 4;

Qy 2 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVFDRLKVDAPKTF 61
Db 23 LPVGNPSPDLLIDGTIWEAGAGDPCDPCATWCDAISLRAGFYGDYVFDRLKVDAPKTF 82

Qy 62 S-HGAKPTGSA--AANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNINWDRFDFVCTLG 118
Db 83 S-HGAKPTGSA--AANYTTAVDRPNPAYNKHLDHAEWFTNAGFIALNINWDRFDFVCTLG 142

Qy 119 SNGYIRGNSTAFNLVGLFGVKGTTVNAVNLPNVSLNGVVELYDTDSFSVSGARGALWE 178
Db 143 SNGYIRGNSTAFNLVGLFGVKGTTVNAVNLPNVSLNGVVELYDTDSFSVSGARGALWE 201

Qy 179 CGCATLGAEFYQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTSKAT 236
Db 202 CGCATLGAEFYQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTSKAT 261

```

Query Match	75.3%	Score 1469.5	DB 2:	Length 380;
Best Local Similarity	71.4%	Pred. No. 2.4e-111;		
Matches 262;	Conservative 54;	Mismatches 48;	Indels 3;	Gaps
Qy	2	LPVGNPSDPSLLDGTWEEAAGPCDPCATWCDAISLRAGFYGDYVDFRILKVDAPKTF	61	
Db	16	LPVGNPAEPSLLDGTWEEGSGDPCDPCATWCDAISLRVFGYGDYVDFRVLKTDYSKMF	75	
Qy	62	SMGAKPTG - SAAANYTTAVDRPNPAYNKKHLHDEWFTNAGFIALNIWDRDVFECTIGASN	120	
Db	76	LMGTAPTSNPSAADSNTAERANPAYGKHMHDEWFTNAGYIALNIWDRDVFECTIGATS	135	
Qy	121	GYIRGNSTAFNLVGLFGVKGTFTVNANELPNVLSISNGVVLYLTTDSFSWSVGARGALWECG	180	
Db	136	GYFGKGNASFNILGIGISGTLTD-QKYPNASTISNGVVLYLTTDSFSWSVGARGALWECG	194	
Qy	181	CATLGAEFQYQAQSKPKVVEELNVICNVISQFVSNNKPKGYKGVAFPLPTDAGVATATGKTSAT	240	
Db	195	CATLGAEFQYQAQSKPRVQELNVLSNVAFTVHKPQGVYGQSLPLPTNAGTSNADLNKAT	254	

Query Match	72.9%	Score 1423;	DB 2;	Length 391;
Best Local Similarity	71.1%	Pred. No. 1.5e-107;		
Matches	266;	Conservative 44;	Mismatches 56;	Indels 8;
Gaps				
Qy	2	LPVGNPDPSSLIDGTWEEGAAGDCDPCATWCDAISLRAGFYGDYVDFRDLTKVDPKTF 61		
Db	18	LPVGNPAEPLSDGTWEEGAGDCDPCATWCDAISLRAGFYGDYVDFRDLTKVDPKTF 77		
Qy	62	S-MGAKPTGSA--AANYTTAVDRENPAYNKHLDHAEWFTNAGFTALNIWDRFDVFCFLGA 118		
Db	78	SGMAAIPTESSGTVSSAKQAVDRVNLAYGKHLQDAEWFTNSAPLALNIWDRFDICTLGA 133		
Qy	119	SNGYIRGNSTAFNLVGLFGYKGT--VNA--NELPNVLSLNGCVVELYTDTSFSWSVGARGA 175		
Db	138	SNGYFKGSSAFLNLVGLFGTAGNESNALNDQLEPNVAITQIVEFYTDTSFSWSVGARGA 197		
Qy	176	LWECGCATLGAEEFOYAQSKPKVSENLICNVSQFSNAKPKGYKGA--FLEPTDAGVATA 233		
Db	198	LWECGCATLGAEEFOYAQSNPKIEMLVNTSSPAQFVIHKPRYKGTTSNFPPLPTAGTDTA 253		
Qy	234	TGTSKATINYHEWQVAGSLSYRLNSLPYIQVQWSRATFDADNTRIAQPKLPTAVLNITA 293		
Db	258	TDTSKATIKYHEWQVGLASYRLNMLVPYIGVNNWSRATFDADTTRIAQPKLATAVLDAKT 313		
Qy	294	WNPSLLGNATLSLTDTSFSDPMQVLSQCIQNKFSRKACGVTVGATLVLDADKWSLTAEARL 353		
Db	318	WNPIITGASGVDTNKNWSDNLQIASIQINKMSRKACGVAVGATLLDADKWSITGEARL 373		
Qy	354	INERAAHVSGQFRF 367		
Db	378	INERAAHMAQFRF 391		

```
RESULT 13
Q9X17 PRELIMINARY; PRT; 341 AA.
AC Q9X17;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydomophila abortus.
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI_TaxID=83555;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=IW508;
RX MEDLINE=93123168; PubMed=8419295;
RA Kalltenboeck B., Kousoulas K.G., Storz J.;
RT "Structures of and allelic diversity and relationships among the major
RL outer membrane protein (ompA) genes of the four chlamydial species.";
DR EMBL: M73040; AAD29103.1;
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam: PF01308; Chlamydia_OMP; 1.
DR ProDom: PD001717; Chlamydia_OMP; 1.
FT NON_TER 1
FT CHAIN 341
SQ SEQUENCE 341 AA; 36762 MW; B5933C9BF6AAFL71 CRC64;

Query Match 72.8%; Score 1420.5; DB 2; Length 341;
Best Local Similarity 75.4%; Pred. No. 1.9e-107;
Matches 257; Conservative 35; Mismatches 48; Indels 1; Gaps 1;

QY 16 GTWEGAAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTFSGAAAN 74
Db 1 GTWEGAAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDVKNKTITGMAVPTGAAAN 60

QY 75 YTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFVCTLGASNGYIRGNSTAFNLVG 134
Db 61 YKTPTRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFVCTLGASNGYIRGNSTAFNLVG 120

QY 135 LFCVKGTNNANLPLNYSLSGVVELYDTSTFSWSVGARGALWECCCATLGAEFFQYQSK 194
Db 121 LGVKGSSAADQLPNWGTQGVFEYDTSTFSWSVGARGALWECCCATLGAEFFQYQSN 180

QY 195 PKVEELNVICNVSVQSVNPKGVAFPLPTDAGVATATGKTSATINHEWQVGSLSY 254
Db 181 PKTEMLNVVSSPAQFVVRHPRGKGTAFPLPTAGTDQATDYSATIKYHEWQVGLALSY 240

QY 255 RLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATALTSTDSFDF 314
Db 241 RLNLVPIYIGVQNSRATFDADNIRIAQPKLAAVNLTAWNPSSLGNATALTSTDSFDF 300

QY 315 MQIVSCQINKSKRACGVTVGATLVDAKWSLTAEARLIN 355
Db 301 LQIASIQINKSKRACGVTVGATLVDAKWSLTAEARLIN 341

RESULT 14
Q46193 PRELIMINARY; PRT; 402 AA.
AC Q46193;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR.
GN OMP.
OS Chlamydia psittaci (Chlamydomophila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
```

```
RX MEDLINE=94103736; PubMed=8277245;
RA Storey C., Lusher M., Yates P., Richmond S.;
RT "Evidence for Chlamydia pneumoniae of non-human origin.";
RL J. Gen. Microbiol. 139:2621-2626(1993).
DR EMBL: L04980; AAA17396.1;
DR InterPro: IPR000604; Chlamydia_OMP.
DR Pfam: PF01308; Chlamydia_OMP; 1.
DR PRINTS: PR01334; CHLAMIDIAOMP.
DR ProDom: PD001717; Chlamydia_OMP; 1.
FT SIGNAL 1
FT CHAIN 23
FT CHAIN 402
SQ SEQUENCE 402 AA; 43261 MW; E36ABC5AF04820A CRC64;

Query Match 72.7%; Score 1419; DB 2; Length 402;
Best Local Similarity 70.5%; Pred. No. 3.2e-107;
Matches 268; Conservative 37; Mismatches 61; Indels 14; Gaps 6;

QY 2 LPVGNPSDPSLLIDGTWEGAAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDAPKTF 61
Db 23 LPVGNPAEPSLLIDGTWEGAAGDPCDPCATWCDALSLRAGFYGVDFVDRILKVDVKNKT 82

QY 62 S-MGAKP---TGSAA--ANYTTAVDRPNPAYNKHLDHAEFTNAGFIALNIWDRFVDFVCT 116
Db 83 SGMAATPTQATGNASNTNQPEANGRNIAAYGRHMDAEWFSNAFLALNIWDRFDFCT 142

QY 117 GASNGYIRGNSTAFNLVGLFGVKGITVNAE----LPNVLSNGVVELYDTSTFSWSVGA 172
Db 143 GASNGYFKSSAAFLNLVGLIGFSSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTSTST 202

QY 173 RGALWECCCATLGAEFFQYQSKRACGVTVGATLVDAKWSLTAEARLIN 230
Db 203 RGALWECCCATLGAEFFQYQSKRACGVTVGATLVDAKWSLTAEARLIN 262

QY 231 ATATGKTSATINHEWQVGSLSYRLNSLPYIGVQNSRATFDADNIRIAQPKLPTAVLN 290
Db 263 TTTATGKTSATIKYHEWQVGLALSRLNLVPIYIGVQNSRATFDADNIRIAQPKLSEIN 322

QY 291 LTAWNPSSLGNATLSTT---DSFDEMOIVSCQINKSKRACGVTVGATLVDAKWSL 347
Db 323 ITWNPSSLGCTALPNNAKGVLSVQIASIQINKSKRACGVTVGATLVDAKWSI 382

QY 348 TAEARLINERAAHVSGQFRF 367
Db 383 TGEARLINERAAHVMAQFRF 402

RESULT 15
Q9AIJ9 PRELIMINARY; PRT; 382 AA.
AC Q9AIJ9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE MAJOR OUTER MEMBRANE PROTEIN PRECURSOR (FRAGMENT).
GN OMPA.
OS Chlamydia psittaci (Chlamydomophila psittaci).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
OX NCBI_TaxID=83554;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MENINGOPNEUMONITIS, MN, ATCC VR122;
RX MEDLINE=2107860; PubMed=11211261;
RA Bush R.M., Everett K.D.;
RT "Molecular evolution of the Chlamydiaceae.";
RL Int. J. Syst. Evol. Microbiol. 51:203-220(2001).
DR EMBL: AF269262; AAK00243.1;
KW Signal.
FT NON_TER 1
FT SIGNAL 2
FT CHAIN 382
SQ SEQUENCE 382 AA; 41231 MW; 691717IA9A69303B CRC64;
```

```

Query Match      72.5%; Score 1415; DB 2; Length 382;
Best Local Similarity 70.3%; Pred. No. 6.3e-107;
Matches 267; Conservative 37; Mismatches 62; Indels 14; Gaps 6;

QY 2 LPVGNPDPSSLIDGTIWEAGAGDPCDPCATWCDAISIRAGFYGDYVDFDRILKVDAPKTF 61
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 3 LPVGNPAEPSSLIDGTIWEAGAGDPCDPCATWCDAISIRAGFYGDYVDFDRVLKVDVNTF 62
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 62 S-MGAKP---TGSAA-ANYTTAVDRPNPAYNKKHLHDAEFNAGFIALNIWDRDFVCTL 116
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 63 SCMAATPTQATGNASNTNQPENGRPNITAYGRHMODAEWFSNAAPLALNIWDRDFICTL 122
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 117 GASNGYIRCNSTAFNLVGLFGVKGTVNANE---LPNVSLNGVVVELYTDTSFSWSYGA 172
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 123 GASNGYFKSSSAFNLVGLIGFSATSTSTELPMOLPNVGTQGVVEFTDTSFSWSYGA 182
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 173 RGALWECGCATLGAEFQYQAQSKPKVEELNVICNVSQFSVKNPKYKGVA--PPLPTDAGV 230
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 183 RGALWECGCATLGAEFQYQAQSNPKIEVLNVTSSPAQFVIHKPRGYKGASSNEPLPITAGT 242
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 231 ATATGTSATINYHEWQVGLASLSYRLNSLVPIGVQWSRATFDADNIRIAQPKLPTAVLN 290
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 243 TEATDTKSATIKYHEWQVGLASLSYRLNMLVPYIGVNWSRATFDADTIRIAQPKLSEILN 302
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 291 LTAWNPSLLGNATALSTT--DSFSDFMQIVSCQINKEKSKACGVTVCATLVDADKWSL 347
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 303 ITTWNPSLLGSTTTLPNNGGKDVLSDLVLIQIASIQINKMKSKACGVAVGATLIDADKWSI 362
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

QY 348 TAEARLINERAAHVSGQFRF 367
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
Db 363 TGEARLINERAAHMAQFRF 382
   |||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||

```

Search completed: February 7, 2002, 21:41:22
Job time: 4651 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using .sw model

Run On: February 7, 2002, 15:48:41 ; Search time 137.02 Seconds
(without alignments)
117.311 Million cell updates/sec

Title: US-09-391-606-7
Perfect score: 1166
Sequence: 1 MTKKHVAVVVEGILNRLPKQ.....DKLGSDFTFRKFDLGIISAF 217

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5
Searched: 522463 seqs, 74073290 residues
Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1162	99.7	422	21	AA192716
2	1162	99.7	438	20	AA193539
3	765.5	65.7	372	20	AA193750
4	88	7.5	951	20	AA194536
5	88	7.5	953	20	AA194403
6	83	7.1	711	19	AA1955103
7	83	7.1	2234	21	AA1981502
8	81	6.9	1091	21	AA1943631
9	80	6.9	303	22	AA1982023
10	80	6.9	359	22	AA1982824
11	79	6.8	327	20	AA197217

12	78	6.7	753	21	AA196159
13	77.5	6.6	2001	18	AA193596
14	77.5	6.6	3672	18	AA191950
15	77.5	6.6	3801	18	AA191949
16	77	6.6	500	19	AA193957
17	76.5	6.6	367	20	AA194087
18	76.5	6.6	367	21	AA198457
19	76.5	6.6	725	20	AA198832
20	75.5	6.5	431	20	AA195534
21	75.5	6.5	725	20	AA198831
22	75	6.4	1329	21	AA191779
23	74.5	6.4	241	19	AA195085
24	74.5	6.4	652	22	AA196046
25	74.5	6.4	730	22	AA196045
26	74.5	6.4	761	21	AA194084
27	74.5	6.4	761	21	AA199418
28	74.5	6.4	761	22	AA196043
29	74.5	6.4	761	22	AA196043
30	73.5	6.3	988	13	AA194302
31	73	6.3	415	21	AA191576
32	73	6.3	476	21	AA1930604
33	72.5	6.2	459	20	AA196942
34	72.5	6.2	561	18	AA198311
35	72.5	6.2	608	22	AA194640
36	72.5	6.2	762	21	AA1937984
37	72	6.2	405	22	AA19943
38	72	6.2	701	22	AA192442
39	71.5	6.1	930	19	AA195961
40	71.5	6.1	1395	20	AA193563
41	71.5	6.1	1395	20	AA198401
42	71	6.1	288	22	AA193621
43	71	6.1	354	21	AA1929125
44	70.5	6.0	112	21	AA192299
45	70.5	6.0	112	21	AA1945000

ALIGNMENTS

RESULT 1
AA192716
ID AA192716 standard; Protein: 422 AA.
AC AA192716;
XX 29-AUG-2000 (first entry)
DT Chlamydia pneumoniae lorf2.
DE Lorf2; vaccine; antibacterial; antigen.
KW Chlamydia pneumoniae.
OS Chlamydia pneumoniae.
PN WO200024901-A1.
XX 04-MAY-2000.
XX 28-OCT-1999; 99WO-GB03565.
PR 28-OCT-1998; 98US-0106037.
PR 20-SEP-1999; 99US-0154658.
PR 26-OCT-1999; 99US-0427501.
XX (CONN-) CONNAUGHT LAB LTD.
PI Murdin AD, Oomen RP, Dunn PL;
XX WPI; 2000-350742/30.
DR N-PSDB; AAA28411.
XX Isolated polynucleotide encoding a Chlamydia polypeptide useful to treat, diagnose and prevent disease caused by Chlamydia infection

Claim 6; Fig 1A-C; 88pp; English.

This is the lorf2 protein of a strain of Chlamydia pneumoniae. Comparison of this sequence as to the recently published genome sequence of C. pneumoniae reveals that the sequence actually contains at least two open reading frames, a first one in the 5' portion and a second one in the 3' portion of the sequence. Despite the presence of the stop codon at the end of this sequence, C. pneumoniae does make a 76 kDa product. It appears possible that C. pneumoniae is able to read through this stop codon and produce a full-length product terminated by the stop codon at the end of the second open reading frame. There is at least one in-frame ATG upstream of the start codon. This suggests that the first open reading frame may form part of one or more larger open reading frames. The lorf2 protein or DNA can be used as a vaccine for humans to treat or prevent disease caused by Chlamydia infection. The sequences or an antibody to lorf2 can be used to diagnose a Chlamydial infection.

Sequence 422 AA;

Query Match 99.7%; Score 1162; DB 21; Length 422;
Best Local Similarity 99.5%; Pred. No. 4.8e-119;
Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 60
DB 206 mtkkhiyawvvegilnrlpkqffvkcsvvdwntfvpsetsttekaatnamkykvcwqwl 265
QY 61 GKHSQVPWINGOKKPLYLYGAFLMNPLAKATKTTLNGKENLAWFTGGTLGRLKAGDWSA 120
DB 266 gkhsqvpwngokkplylygafilmplakatkttlngkenlawfignlgrkagdwsa 325
QY 121 TVRYEYVEALSVPEIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180
DB 326 tvryeyvealsvpelidvsgigrgnllkfwfaqaiaanydpkeangftnykgsalymygi 385
QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217
DB 386 tdslsfraygayskpandksgdftfrkfdlgiisaf 422

RESULT 2

AA35359
ID AAY35359 standard; Protein; 438 AA.

XX AAY35359;

XX 13-SEP-1999 (first entry)

XX Chlamydia pneumoniae transmembrane protein sequence.

XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.

XX Chlamydia pneumoniae.

XX WO9927105-A2.

XX 03-JUN-1999.

XX 20-NOV-1998; 98WO-IB01890.

XX 04-NOV-1998; 98US-0107078.

XX 21-NOV-1997; 97FR-0014673.

XX (GEST) GENSET.

XX Griffais R;

XX WPI; 1999-357842/30.

XX Genome sequence of Chlamydia pneumoniae

XX

PS Page 1157; Disclosure; 1912pp; English.

XX AAY34584-Y35879 represent the proteins encoded by all the open reading
CC frames in the complete genome (see AAY34584-Y35879) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.

SQ Sequence 438 AA;

Query Match 99.7%; Score 1162; DB 20; Length 438;
Best Local Similarity 99.5%; Pred. No. 5.1e-119;
Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVDWNTFVPSETSTTEKAATNAMKYKVCWQWLV 60
DB 222 mtkkhiyawvvegilnrlpkqffvkcsvvdwntfvpsetsttekaatnamkykvcwqwl 281
QY 61 GKHSQVPWINGOKKPLYLYGAFLMNPLAKATKTTLNGKENLAWFTGGTLGRLKAGDWSA 120
DB 282 gkhsqvpwngokkplylygafilmplakatkttlngkenlawfignlgrkagdwsa 341
QY 121 TVRYEYVEALSVPEIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYMYGI 180
DB 342 tvryeyvealsvpelidvsgigrgnllkfwfaqaiaanydpkeangftnykgsalymygi 401
QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217
DB 402 tdslsfraygayskpandksgdftfrkfdlgiisaf 438

RESULT 3

AA37570
ID AAY37570 standard; Protein; 372 AA.

XX AAY37570;

XX 07-OCT-1999 (first entry)

XX Chlamydia trachomatis cellular envelope protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perinephritis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX
PS
XX
CC
CC
CC
CC
CC
CC
CC
CC
CC
CC
XX
SQ

Disclosure: Page 1226; 1755pp; English.

AAV36754-Y37949 are encoded by open reading frames (ORFs) of the genome of Chlamydia trachomatis (see AA01425). The polypeptides can be used as vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences can also be used to control growth of the microorganism. Chlamydia trachomatis is responsible for a large number of diseases, e.g. eye diseases such as conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion conjunctivitis; genital diseases such as nongonococcal urethritis, epididymitis, cervicitis, salpingitis, perihepatitis, Bartholinitis; pneumopathy in breast feeding infants; and venereal lymphogranulomatosis. The polypeptides of the invention may be of use in treating these diseases.

SQ Sequence 372 AA;

Query Match 65.7%; Score 765.5; DB 20; Length 372;

Best Local Similarity 62.8%; Pred. No. 1.4e-75;

Matches 147; Conservative 27; Mismatches 41; Indels 19; Gaps 4;

QY 1 MTKKHVAVVEGILNRLKQFVKGVVDWNTFV-----PSETSTPEKAATNAMKYKCV 55
Db 141 maekeyawveailnklpfnfvvktsvldwntltaktndpadastagpakpnt-kydyly 199
QY 56 WQWLYGKHSQVFWINGQKPLLYLYGAFLMNPLAK-----ATKTTINGKRENLA 103
Db 200 wqwlvgkstamwpfngqtknlytygaylfnplaelpenwkqstttatpkit-ngkenhaw 258
QY 104 FIGTGLGLRAGDSATVREYVEALSVPEIDVSGIGRNLKFWFAQAIAANYDPKEA 163
Db 259 figcslgvrragdwsatvreyvealpaelpdvagigrngmqkygfaaikqgidpkes 318
QY 164 NSFTNYKGFSAIYMYGITDLSFRAYGAYSKPANDKLGSDFTFERFDLGIISAF 217
Db 319 ngftnykxvsyqfvmgltdsvsfrrayayskpandnlgdsdftcyrydydglissf 372

RESULT 4

AAV34536
ID AAV34536 standard; Protein; 951 AA.

AC AAV34536;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG67.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic.

OS Porphyromonas gingivalis.

PN W09929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;

XX WPI; 1999-385613/32.

DR IN-PSDB: AAX91754.

XX Antigenic Porphyromonas gingivalis peptides for preventing

PT gingivitis

XX Claim 1; Page 526-527; 588pp; English.

XX AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAX34318 to
CC AAX34583. AAX91802 to AAX91999 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.

XX Sequence 951 AA;

Query Match 7.5%; Score 88; DB 20; Length 951;

Best Local Similarity 30.1%; Pred. No. 1.4;

Matches 25; Conservative 17; Mismatches 35; Indels 6; Gaps 4;

QY 117 DWSATVREYVEALSVPEIDVSGIGRG-NLLKFWFAQAIAANYDPKEAN---SFTNYKGF 172

Db 495 dwkngmrhsvplstvtplldylnltmgvnynewytkgrkswnedkktflpsdttyk-f 553

QY 173 SALYMYGITDLSFRAYGAYSKP 195

Db 554 rrllydyslsaglsttlygmf-kp 575

RESULT 5

AAV34403
ID AAV34403 standard; Protein; 953 AA.

AC AAV34403;

DT 25-AUG-1999 (first entry)

DE Porphyromonas gingivalis protein PG67.

KW Porphyromonas gingivalis; PG; periodontal disease; gingivitis;

KW vaccine; antigenic.

OS Porphyromonas gingivalis.

PN W09929870-A1.

PD 17-JUN-1999.

PF 10-DEC-1998; 98WO-AU01023.

PR 04-AUG-1998; 98AU-0005028.

PR 10-DEC-1997; 97AU-0000839.

PR 31-DEC-1997; 97AU-0001182.

PR 30-JAN-1998; 98AU-0001546.

PR 10-MAR-1998; 98AU-0002264.

PR 09-APR-1998; 98AU-0002911.

PR 23-APR-1998; 98AU-0003128.

PR 05-MAY-1998; 98AU-0003338.

PR 22-MAY-1998; 98AU-0003654.

PR 29-JUL-1998; 98AU-0004917.

XX (CSLC-) CSL LTD.

PI Agius CT, Barr IG, Hocking DM, Margetts MB, Patterson MA;
PI Ross BC, Rothel LJ, Webb EA;

XX WPI: 1999-385613/32.
DR N-PSDB; AAX91621.
XX Antigenic Porphyromonas gingivalis peptides for preventing
PT gingivitis
XX
PS Claim 1; Page 373-375; 588pp; English.
XX
CC AAX91536 to AAX91801 encode two hundred and sixty six antigenic
CC Porphyromonas gingivalis (PG) polypeptide sequences given in AAY34318 to
CC AAY34583. AAX91802 to AAX91989 represent PCR primers used in the
CC isolation of the PG polypeptides. The PG polypeptides have antibacterial
CC activity with a vaccine mechanism of action. The PG polypeptides can be
CC used as vaccines especially against Porphyromonas gingivalis. Probes can
CC be used to detect Porphyromonas gingivalis in standard hybridisation
CC assays. Porphyromonas gingivalis is involved in periodontal disease
CC especially gingivitis.
XX
XX Sequence 953 AA;
SQ

Query Match 7.5%; Score 88; DB 20; Length 953;
Best Local Similarity 30.1%; Pred. No. 1.4;
Matches 25; Conservative 17; Mismatches 35; Indels 6; Gaps 4;
QY 117 DMSATRVRYEYVVALSVPEIDVSGIGRG-NLLKFWFAQIAAANYDPKEAN---SFTNYKGF 172
DB 497 dwkngmrhvspisltpvllldylnltmgvnynewytkgrkswnedkktflpsdtyk-f 555
QY 173 SALLYXGITDSLSFRAYGAYSKP 195
DB 556 rrllydsiaagisttlygmf-kp 577
RESULT 6
ID AAW55103 standard; Protein; 711 AA.
XX
AC AAW55103;
DT
DT 02-OCT-1998 (first entry)
XX Streptococcus pneumoniae SP0051 protein.
DE
XX Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;
KW detection; pneumonia; otitis media; meningitis.
XX Streptococcus pneumoniae.
OS
PH Key Location/Qualifiers
FT Misc-difference 56 /label= unknown
FT /note= "encoded by NAG"
XX
XX WO9818930-A2.
XX
XX 07-MAY-1998.
XX
XX 30-OCT-1997; 97WO-US19422.
XX
XX 31-OCT-1996; 96US-0029960.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Choi GH, Hromockyj A, Johnson LS, Kunsch CA;
XX
XX WPI: 1998-272224/24.
DR N-PSDB; AAV27364.
XX
XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
PT pneumoniae - or their epitope-containing fragments, useful in
PT protective or therapeutic vaccines, and for diagnosis
PT

XX Claim 11; Page 66; 118pp; English.
XX
CC The present sequence represents a protein from Streptococcus pneumoniae.
CC The nucleic acid sequence encoding the Streptococcus pneumoniae protein
CC can be useful in vaccines for inducing protective antibodies against
CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
CC are used to detect Streptococcus infection (by usual hybridisation or
CC amplification methods), also for isolating Streptococcus genes or their
CC allelic variants. The protein can be used similarly to detect specific
CC antibodies in standard immunoassays, especially for diagnosing or
CC monitoring infections. Antibodies which bind the protein are used to
CC detect corresponding antigens, to purify the protein and for passive
CC immunisation (optionally coupled to a toxin). Vaccines are administered,
CC e.g. by injection, orally or through the skin, typically at 0.01-1000
CC (especially 10-300) mu g/ml per dose.
XX
XX Sequence 711 AA;
SQ

Query Match 7.1%; Score 83; DB 19; Length 711;
Best Local Similarity 23.4%; Pred. No. 3.2;
Matches 57; Conservative 28; Mismatches 81; Indels 78; Gaps 14;
QY 32 TFPVSETSTTEKAATNAMYKYCVWOLYKHSQVPIWGQKPKLYLYGAFI---MNPL 87
DB 443 thnpaseqtliqaaeigl-----lvqeeafdtwygg-kkp-ydygrffekdathpe 491
QY 88 AKA-----TKTTL-NGKENLA---WFITGTLGLRKAGDWSATRVRYEYVALSVPEI 135
DB 492 arkgekwsdfdlrtmvergknpafimwsigneigangdahslatvk-rlvkv--ikdv 548
QY 136 DVSGIGRGNLLKFWEA-----QAIYAANYDPKEANSFTNYKGFSAIY----MY 178
DB 549 dktryvtmgadkfrfgngsgghekiadelavgfny-----sednykalrakhpkwliy 602
QY 179 GITDSLSFRAYGAYSKP-----ANDKLG-----SDFTFKFDLGI 213
DB 603 gsetssatrtgryrperelkhsngpernyeqsdgndrvwgkktaswtfdnag 662
QY 214 ISAF 217
DB 663 agqf 666
RESULT 7
ID AAY81502 standard; Protein; 2234 AA.
XX
AC AAY81502;
DT
DT 24-MAY-2000 (first entry)
XX Streptococcus pneumoniae type 4 protein sequence #2.
DE
XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.
XX Streptococcus pneumoniae.
OS
XX WO200006737-A2.
XX
XX 10-FEB-2000.
XX
XX 27-JUL-1999; 99WO-GB02451.
XX
XX 27-JUL-1998; 98GB-0016337.
PR 19-MAR-1999; 99US-0125164.
XX
XX (MICR-) MICROBIAL TECHNIQS LTD.
PA
XX

PI Gilbert CFC, Hansbro PM;
 XX WPI: 2000-195300/17.
 XX New Streptococcal protein, useful as a vaccine, for diagnosis of
 PT pneumococcal diseases and for screening agents capable of antagonizing
 PT or inhibiting expression of the protein
 XX
 PS Claim 1; Page 72; 108pp; English.
 XX
 CC AAY81501 to AAY81679 represent specifically claimed protein sequences
 CC isolated from Streptococcus pneumoniae. AAA05407 to AAA05590 represent
 CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
 CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonizing, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. AAA05591 to AAA05614 represent primers used in the
 CC exemplification of the present invention.
 XX
 SQ Sequence 2234 AA;

Query Match 7.1%; Score 83; DB 21; Length 2234;
 Best Local Similarity 23.4%; Pred. No. 16;
 Matches 57; Conservative 28; Mismatches 81; Indels 78; Gaps 14;
 QY 32 TFPVSTETTERKAATNAMYKYCYVQWLVGKHSQVPWINGOKKPLLYXGAFI----MNPL 87
 Db 483 thnpaseqtlqiaaelj-----lvqeeafdtwygg-kkp-ydygrffekdathpe 531
 QY 88 AKA-----TKTTL-NGKENLA---WFIGTGLGLRKGADWSATVRYEYVEALSVPDI 135
 Db 532 arkgewksdfdlrtmvergknpafmwsigneigangdahslatvk-rivkv--ikdv 588
 QY 136 DVSIGIGRGNLLKFWFA-----QAIAANYDPKEANSFNYKGFSAIY----MY 178
 Db 589 dktrytngadfrfngsgghekiadelavgfvy-----sednykairakhpkiwly 642
 QY 179 GTDLSLFRAYGAYSKP-----ANDKLG-----SDFTFRKFDLGI 213
 Db 643 gsetssatrtgsvyirperelkhsngpernyeqsdgndrvywgktataswffdrdnagy 702
 QY 214 ISAF 217
 Db 703 agqf 706

RESULT 8
 AAY43631
 ID AAY43631 standard; Protein; 1091 AA.
 XX
 AC AAY43631;
 XX
 DT 11-FEB-2000 (first entry)
 XX
 DE Amino acid sequence of the HMG-CoA synthase enzyme.
 XX hmc gene; 3-hydroxy-3-methylglutaryl-CoA synthase; HMG-CoA synthase;
 KW mevalonate pathway; carotenogenic yeast; isopentenyl pyrophosphate;
 KW farnesyl pyrophosphate; isoprenoid; carotenoid; astaxanthin;
 KW cancer; antioxidant; colouring reagent; farmed fish industry.
 XX
 OS Phaffia rhodozyma.
 XX
 PN EP955363-A2.
 XX
 PD 10-NOV-1999.
 XX

PF 26-APR-1999; 99EP-0107413.
 XX
 PR 06-MAY-1998; 98EP-0108210.
 XX
 FA (HOFF) HOFFMANN LA ROCHE & CO AG F.
 XX
 PI Hoshino T, Ojima K, Setoguchi Y;
 XX
 DR WPI: 2000-001086/01.
 DR N-PSDB; AAZ30171.
 XX
 PT Isolated DNA sequences encoding enzymes, useful for the production of
 PT isoprenoids and carotenoids
 XX
 PS Claim 3; Page 37-40; 58pp; English.
 XX
 CC The present sequence is represents a 3-hydroxy-3-methylglutaryl-CoA
 CC synthase (HMG-CoA synthase) enzyme, and is encoded by the hmc gene.
 CC The enzyme is involved in the mevalonate pathway in the carotenogenic
 CC yeast Phaffia rhodozyma. The specification also describes enzymes that
 CC are involved in the pathway from isopentenyl pyrophosphate to farnesyl
 CC pyrophosphate. The enzymes of the invention are used in the production
 CC of isoprenoids and carotenoids, especially astaxanthin. Astaxanthin is
 CC useful for the pharmaceutical industry, to protect cells against cancer
 CC as it has a strong antioxidant property. Astaxanthin is also useful
 CC as a colouring reagent in the farmed fish industry, e.g. salmon.
 XX
 SQ Sequence 1091 AA;
 Query Match 6.9%; Score 81; DB 21; Length 1091;
 Best Local Similarity 22.4%; Pred. No. 9.8;
 Matches 48; Conservative 30; Mismatches 50; Indels 86; Gaps 12;
 QY 8 WV-VEGI-----LNRLPKQFFVKCSVVDWNTFPVSTET-----TEKA 44
 Db 756 wldsevegmevmaasfnstsfarlsqskcmagrslyrlatstgdangmmagkteka 815
 QY 45 ATNAMY-----KYCVQWLVGKHSQVPWINGOKKPL-----VLYGAFI-MNPLAK 89
 Db 816 letiseyfpmsqilalsngycidk----kpsainwiegrkvsvaesvipgaivksvl-- 869
 QY 90 ATKTT-----LNKENLAWFIG-----CTLGGLR-----KAGDWSATV 122
 Db 870 --kttvadlvlnniknl---igsamagsiggnfnahasdiltstflgtgqdpagqvessm 924
 QY 123 RYEYVEAL-----SVPEIDVSGIGRGNLL 146
 Db 925 cmtmeavndgkdllitcsmpalecvtvggtfl 958
 RESULT 9
 AAG82023
 ID AAG82023 standard; Protein; 303 AA.
 XX
 AC AAG82023;
 XX
 DT 03-SEP-2001 (first entry)
 XX
 DE S. epidermidis open reading frame protein sequence SEQ ID NO:1140.
 XX
 KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
 KW vaccination; endocarditis.
 XX
 OS Staphylococcus epidermidis.
 XX
 PN WO200134809-A2.
 XX
 PD 17-MAY-2001.
 XX
 PF 09-NOV-2000; 2000WO-US30782.
 XX
 PR 09-NOV-1999; 99US-0164258.
 XX

WP	WO200134809-A2.
PD	17-MAY-2001.
XX	
XX	09-NOV-2000; 2000WO-US30782.
XX	
XX	09-NOV-1999; 99US-0164258.
PPR	(GLAX) GLAXO GROUP-LTD.
XX	Kimmerly WJ;
XX	WPI; 2001-316495/33.
XX	N-PSDB; AAH53674.
XX	
PPT	Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
PPT	useful for vaccinating against infections, e.g. endocarditis -
XX	
XX	Claim 18; Page-714; 2188pp; English.
XX	
CC	AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
CC	(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
CC	(I) and (II) can have antibacterial activity and therefore can be used
CC	in vaccination. The nucleic acids (I) may be used to produce the
CC	S. epidermidis polypeptides (II) via the production of vectors
CC	containing them which are used to produce hosts cells which express the
CC	polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
CC	used to vaccinate subjects and to raise antibodies against the bacteria.
CC	The polypeptides may also be used to assay for other inhibitors of their
CC	activity and therefore identify compounds that may be used for the
CC	treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
CC	AAH55030 represent specifically claimed S. epidermidis genomic DNA
CC	polynucleotide sequences from the present invention. AAH55091 to
CC	AAH55098 represent oligonucleotide sequences and primers which are used
CC	in the exemplification of the present invention.
CC	N.B. The present invention specifically claims all the polynucleotide
CC	sequences given in the sequence listing of the present specification,
CC	however the sequence listing only goes up to SEQ ID NO:4454 so even
CC	though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
CC	no sequences are present for SEQ ID NO:4455 to 4464.
XX	
XX	Sequence 359 AA;
SEQ	
	Query Match 6.9%; Score 80; DB 22; Length 359;
	Best Local Similarity 24.6%; Pred. No. 2.7;
	Matches 45; Conservative 22; Mismatches 48; Indels 69; Gaps 12;
QY	67 PWINGOKKPLY-----LYGAFLNNPLAKATKTTLNGKENL-----AWFTG-106
Db	: :
	197 pwvieskttvfesrvpllldnnhyrylgfllh-----qlngkelimitedlwsI- 247
QY	107 GTLGGLRGADSWATRVYEVALESPVIDVGIGRG-----NLLKF-----WFAQ 152
Db	: : : : : : : : : : : : : :
	248 -----lesmndyekilytlvggtlnklidf--ihrgmqrlrnfkfktytsftdwiq 300
QY	153 A---IAANYDPKEANSFTNYGFSGALYMYGITDLSFPRAYGAYSKEPANDK-LGSDTFTRK 208
Db	: : : : : : : : : : : : : :
	301 aemiaetavdlvdvdyv-----aafvy-----lsyr-----rssqpltkrqldmdfnvr 346
QY	209 FDL 211
Db	:
	347 ykI 349
RESULT	11
AAV17217	
ID	AAV17217 standard; Protein: 327 AA.
XX	
AC	AAV17217;
XX	
DT	03-AUG-1999 (first entry)
XX	


```
/transl_table=11
/product=Flagellar Motor Switch Domain/YscQ family"
/protein_id="AAdL8943.1"
/db_xref="GI:4377007"
/translation="MAVAADSSASWIKSRNNFLTSLGTEQVAAPPEPKLQKHIR
EKPIEDVYSIKRGIATAVATGFEFGVHLLIDPMVYQPVENVENLFLTSEEDQLQEL
```

Query Match	99.7%	Score 1098	DB 1	Length 20057
Best Local Similarity	100.0%	Pred. No. 1.2e-275		
Matches 1098	Conservative 0	Mismatches 0	Indels 0	Gaps 0
Qy	4	ttgcctgtagggaacccctctctgacccaagcttattaatgtatggtacaatatggaaggt	63	
Db	408	TTGCCTGTAGGAACCCCTTCTGATCCAAGCTTATTAAATGATGTTACAAATATGGAAGGT	467	
Qy	64	gctcaggagatccttgogataccttgogctacttggctgcagcgtctattagcttcagtgct	123	
Db	468	GCTCAGGAGATCCTTGGCGATCCTTGGCTACTTGGTGCAGCGCTATTAGTCTACGTGCT	527	
Qy	124	ggatttacgagactatgttttcgacgctatcttaaaagtatgacacctaaacattt	183	
Db	528	GGATTTTACGGAGACTATGTGTTTGACCGTATCTTAAAGTAGATGCACCTTAAACATTT	587	
Qy	184	tctatggagccaagcctactggatccgcctgctgcaaaactactactgcogtagataga	243	
Db	588	TCTATGGAGCCAAAGCCTACTGGATCCGCTGCTGCAAACTACTACTTCCGCTAGATAGA	647	
Qy	244	cctaaccggcctacaataaagcatttacagatgcagagtggttcactaaatgcaggttc	303	
Db	648	CCTAACCCGGCTTACAATTAAGCATTTACCATGCAGAGTGGTTCACTAATGCAGGCTTC	707	
Qy	304	attgccttaaacatttgagatcgcttggatgttttctgactttaggagcttctaaggt	363	
Db	708	ATTGCTTTAAACATTTGGGATCGCTTTGATGTTTCTGACTTTAGGAGCTTCTAATGGT	767	
Qy	364	tacattagagaaactctacagcttcctaactcgtttggttattcoggagttaaaggtact	423	
Db	768	TACATTAGAGAAACTCTACAGCGTTCATCTCGTTGGTTTATTCCGGAGTTAAAGGTACT	827	
Qy	424	actgtaaatgcaatgaactaccaacgtttcttaagttaacgagtggttgaacttaac	483	
Db	828	ACTGTAAATGCAAAATGAACACTCCAACCGTTCTTTTAAGTAACGGAGTTGTGCACTTTAC	887	
Qy	484	acagacacctcttcttggagcgtgagcgtctgagccttatgggaatcgcggtgt	543	
Db	888	ACAGACACCTCTTCTTGGAGCGTAGGCGCTGCTGGAGCCTTATGGGAATCGCGTTGT	947	
Qy	544	gcaacttggagcgtgaattccaatgacacagtcacaaacctaaagtgtgaagaacttaac	603	
Db	948	GCAACTTTGGGAGCTGAATTCCAATATGCACAGTCCAAACCTTAAAGTTCAAGAACTTAAT	1007	
Qy	604	gtgactgtaacgtatcgcaattctctgttaacaaacccaagggctataaagcggtgtct	663	
Db	1008	GTGACTGTAAAGTATCGCAATCTCTGTAAACAAACCCCAAGGCTATAAAGCGTTGCT	1067	
Qy	664	ttcccccttgccaacagacgtggcgtagcaacagctactctggcaaaaagtctcgaccatc	723	
Db	1068	TTCCCCCTTGCCAACAGACGCTGGCGTAGCAACAGCTACTTGGAAACAAAGTCTCGACCATC	1127	
Qy	724	aattatcatgaatgcaagttaggcctctctattcagactaaaactcttagtgcca	783	
Db	1128	AATTATCATGAATGGCAAGTAGAGGCTCTCTATCTTACAGACTAAACCTCTTTAGTGCCA	1187	
Qy	784	tacattggagtacaattgtctcgacaacttttgatgctgataaacatccgcatgtgctcag	843	
Db	1188	TACATTGGAGTACAATGGTCTCGAGCACTTTTGTATGCTGATAAACATCCGCATGCTCGAG	1247	
Qy	844	caaaactacctaactgcttttaaaacttaactgcatggaaccctcttacttaggaat	903	
Db	1248	CCAAAACTACCTACAGCTGTTTAAACTTAACATGATGGAACCCCTTCTTTACTTAGGAAT	1307	
Qy	904	gccacagcattgctactactgattcgttctcagacttcgatgataaatgttttctcgtcag	963	

RESULT 5	AE002167	25150 bp	DNA	BCT	06-APR-2001
LOCUS	Chlamydomydia pneumoniae AR39,	section 4 of 94	of the complete		
DEFINITION	genome.				
ACCESSION	AE002167	AE002168	AE002161		
VERSION	AE002167.2	GI:8163349			
KEYWORDS					
SOURCE	Chlamydomydia pneumoniae AR39.				
ORGANISM	Chlamydomydia pneumoniae AR39				
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.				
AUTHORS	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.				
TITLE	Genome sequences of Chlamydia trachomatis Mohn and Chlamydia pneumoniae AR39				
JOURNAL	Nucleic Acids Res.	28	(6),	1397-1406	(2000)
MEDLINE	20150255				
PUBMED	10684935				
REFERENCE	2	(bases 1 to 25150)			
AUTHORS	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (01-MAR-2000)	The Institute for Genomic Research,	9712		
COMMENT	Medical Center Dr. Rockville, MD 20850, USA				
	On Jun 1, 2000 this sequence version replaced gi:7188982				
	gi:7188971.				
FEATURES	Location/Qualifiers				
source	1. .25150				
	/organism="Chlamydomydia pneumoniae AR39"				
	/strain="AR39"				
	/db_xref="taxon:115711"				
	/note="synonym: Chlamydia pneumoniae AR39"				
gene	306. .1325				
	/gene="CP0032"				
CDS	306. .1325				
	/gene="CP0032"				
	/note="similar to GB:M25323 SP:P13580 GB:M30785 GB:X174				
	P1D:146330; identified by sequence similarity; putative				
	/codon_start=1				
	/trans_table=1				
	/product="glucamyl-tRNA reductase, putative"				
	/protein_id="AAF37927.1"				
	/db_xref="gi:7188972"				
	/translation="MVLGVGVSIREAAKERAIQYLOQSEKHLFLAQRFLQK				
	FIPLTETQAGVYKSPKQERELPDLHFLQKALKEGYSRIGFPHQVPT				
	LVQETLLSYDKSIYTNFLVGYSDINRKVAAYLYQHGVTFCSROCVTPAYPT				
	ETLDFRQPDYVIFGSSSEQSDLSIASIPKRIVFDENFVRFLWKEPTP				
	YLDIDFISCEVQKLOCTREGVNRKALLLTCAAKQWEIYEKKSHITQRIQSS				
	SVLSY"				
gene	1692. .2084				

```

/ gene="CP0033"
1692..2084
/ gene="CP0033"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37928.1"
/ db_xref="GI:7188973"
/ translation="MLEKLNFAFYMGITSLDADGAVLPISEVVKVRAQONAD
NEVLSALGALPPSADTAKLYLQMMIGNLFGRETGGALGLDSEGNVVMVRRSGDT
TYDDFVRHVESFMFSETWLSDLGLGQ"
2098..4635
/ gene="CP0034"
2098..4635
/ gene="CP0034"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37929.1"
/ db_xref="GI:7188974"
/ translation="MAVRLVDEGPLSGVIFVLEDIGISWIGRSSANDIPEDPKLG
ASQALINTDGSYITNLDITPIVNGVAIQETTLQKNEDTILGNSQYSLDFEFD
PQDLVYDFDIPENFSDGSDSNGOQKLEPROTSETNHSKPKKELTKDOGSSD
PITSDQQLADAFASAKAEKNOPRAKVAKKGLKESSEINPKQONAKDSPKEERT
NKQNAIMEDGASPRODPOPKSAEPSLKNARTDETLKKNKPVEEKANKKATPDSPE
KDKQPEESKEGSKIETPLDSOKESDEKAEAFVQEEENITENKEDSDSAADA
NDTASDTADENKTPKKNENKSAVLSPPHQDLFRDQTIFFPAEIDDDIAKKNISV
DLTQPSRELLVLGAGANAEHFDLSGKTYILGDTPTCDIVFNDLSVSHQAKITV
NGGILIEDLSKNGVIVEGRKIDKSTLSSNVVVALGTTLLFDHHPADTIVASL
SPDDYSLFGROQDAEALRQAEQEEKOKRATLPAGSFILTLFVGGIALLFGIGTAS
LFLKREVPLENDYDDEAVQEMNILLKRPESKISMSPKPKFLITGYKITEEQ
RACLVDYLNIHFNYSLLENKVVYETQMLKALHGLGOGFANHVAFVNGEVLITGY
VNDQAEFRVAVQELSGIPGRLVKNFAVLLPAEEGILDLNLRPNRYRTVGYSTRY
EISINVNVGRLRGDVIDGMTVTISIQPNAIFLEKGLKYKIDYNK"
4682..4930
/ gene="CP0035"
4682..4930
/ gene="CP0035"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37930.1"
/ db_xref="GI:7188975"
/ translation="MFNMNTAKEKNQSPLLDLEQDMQDHRAQELKASVQDKVHKL
HALLRGSDKESFCQQQSLLAGYVALQVLGRINKMI"
4952..5206
/ gene="CP0036"
4952..5206
/ gene="CP0036"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37931.1"
/ db_xref="GI:7188976"
/ translation="MATNKSCTADFNMKLDGCVTYKGVQOYLTELETSTGQTVDLG
TMFNLFQRMQILSQYMESVSNILTAVENTMITMARVKS"
5225..5674
/ gene="CP0037"
5225..5674
/ gene="CP0037"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11

```

```

/ product="conserved hypothetical protein"
/ protein_id="AAF37932.1"
/ db_xref="GI:7188977"
/ translation="MADLEVFQADFALLFEAGLIAIKQGDSDARKLQSLHILNPNH
YGHDLGALISLHKMDLFAEERLSALIKGNEDNMSKAFSLTHMLTILVHQSSFEV
RRSLESLCKFAQDVIANCKIESTRALAQSVLDWHDITLVAKSAGPLG"
5708..6382
/ gene="CP0038"
5708..6382
/ gene="CP0038"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37933.1"
/ db_xref="GI:7188978"
/ translation="MIDPECFPNLDGAEASQITONSQGTPLASELKKDISPFALGSY
AAPKDTTLVQGFKNPMAMQDQNSNLIDPELQALESELEQEQINNLKRLWDFRST
FEDSOTTAOFADAEHFOAVGVIIDILNEDLNTIAEHTQODARKDEKESGVTKKIIDW
SSGEVLNRALLYFSDRDGNRESLANFLKVQVQVQATQRAELFASIVGTSVSSVKTI
MTTQLG"
6384..7712
/ gene="CP0039"
6384..7712
/ gene="CP0039"
/ note="similar to SP:P21211 GB-U02499 SP:P21212 SP:P40290
PID:437202; identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="type III secretion cytoplasmic ATPase Scn"
/ protein_id="AAF37934.1"
/ db_xref="GI:7188979"
/ translation="MDQLTDFTLMSQLGPNLTTVVGRITEVVGMLIKAVPNVRV
GEVCLVKGMEPLVTEVVGTSQFAELSPGLSGVSPSEVPTGCLPIHRAAGNL
LGRVLNGICEFIDVETGKPLQNVDTFFFRAPDPLHRAKLQILSTGVRICIDMLT
VARQOIGFAGAGVGKSSLLGMIAARNAEADVNIALIGREGRVREFIEGDLGEG
MKRSVYVYSTDSQSLRNAAYTAIAYFRDQGTTVLMDMSVTRFAALREVGL
AAGPPPARAGTSPSVFTSLRLLERSGASDKGTITAFYTVLVAGDMNEPVADEVKSI
LDGHVLSNALAQAHYPAIDVLASISRLLTAVIPEQRRIIGKAREVLAKYKANEML
IRIGEYRGRSDREIDFAIDHIDKLNRLFKQDIEKNTYEEAAQQLAIRF"
7730..8236
/ gene="CP0040"
7730..8236
/ gene="CP0040"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37935.1"
/ db_xref="GI:7188980"
/ translation="MAKYPLEPVLAIKKDRVDRAEKVVVKKRRLLIEBQELRKEAE
RDKVKNHYMQIKQOLRDLDEGTTSDAVLOIKSVIKVAVOLSEEEKVKNQKVVLA
ASKELEKAEVNLAKRREKREKRLHKEEMKEALKEARAEKEQDEMGQLLFLRLRQK
KKRESGGS"
8240..9082
/ gene="CP0041"
8240..9082
/ gene="CP0041"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37936.1"
/ db_xref="GI:7188981"
/ translation="MELAKKTAEISYAKTDNHTVYQNSPEPRDSRDVVFSLQKQTR
QEKTTSSKGNTRTESRPFADEEKRVDDIEAVGSEKEEQSEOECELAENAFAGSLID
IAAAGSAEAVVEVAPIAVVSSITQMIENIILSTVESMVVISINQLVELVLDASSV
PEAFVGANLTLVQSQDLVSRKFSFVDATQAEADLVTNPNPSQLSSLSVALKGHQLT

```

Query Match 99.7% Score 1098; DB 1; Length 25150;

Best Local Similarity 100.0%; Pred. No. 1.2e-275;
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgccttagggaacccctctgatcccaagccttattaattgtagtgcataatggaaggt 63
|||||
Db 21128 TTGCCTTAGGGAACCCCTCTGATCCAAAGCTATTAAATTGATGGTACAAATATGGAAGGT 21069

QY 64 gtcgaggagatcccttgatcccttgctgacttgctgagcgtatttagcttagct 123
|||||
Db 21068 GCTGAGGAGATCCCTTGGGATCCCTTGGCGTACTTGGTGGGACGCTATTAGCTTACGTGCT 21009

QY 124 gqattttacggagactatgttttgcacgctattcttaaaagtagatgcacccctaaacattt 183
|||||
Db 21008 GGAATTTACGGAGACTATGTTTTCACCGTATCTTAAAGTAGATGCACCTAAACATTT 20949

QY 184 ttcttgggagccaaacctactggatccgctgctgcaactactactactgcgctagataga 243
|||||
Db 20948 TCTATGGGAGCCAAAGCTACTGGATCCGCTGCTGCAAACTATACTACTGCGGTAGATAGA 20889

QY 244 cctaaccggcctacaataagcatttacagatgcagagtggttcactaatgcaggcttc 303
|||||
Db 20888 CCTAACCCGGCCTACAATAAGCATTTACAGATGCAGAGTGCTTCACTAATGCAGGCTTC 20829

QY 304 attgccttaaacatttgggagcgttttgatgttttctgtacttttaggagccttctaaggt 363
|||||
Db 20828 ATTGCCTTAACATTTGGGATCGCTTTGATGTTTCTGTACTTTAGGAGCTTCTAATGCT 20769

QY 364 tacattaggaactctacagcgttcactctctgttatttcggaggttaaggtact 423
|||||
Db 20768 TACATTTAGAGAACTCTACAGGTTCAATCTCGTTGGTTTATTTCGGAGTTAAAGGTACT 20709

QY 424 actgtaaatgcaaatgaactacccaacgctttcttaagttaacgaggtgttgaactttac 483
|||||
Db 20708 ACTGTAATGCAAAATGAATCAACCAACGTTCTTTAAGTAACGAGGTGTTGAACCTTAC 20649

QY 484 acagacacctcttctcttgaggcgttagcgctctgtagccttattggaatgcggttgt 543
|||||
Db 20648 ACAGACACCTCTTCTCTTGGAGCGTAGGCGCTGCTGGAGCCTTATGGAATGCGGTTGT 20589

QY 544 gcaactttggagcgtgaattccaaatagcacagtcacaaacctaaagtgaagaactaat 603
|||||
Db 20588 GCAACTTTGGAGCTGAAATTCCAATATGCACAGTCAACACCTTAAAGTTGAAGAATTAAT 20529

QY 604 gtgactgtaacgtagcgaattctctgttaaacacacccaaagcgtataaagcgcttgt 663
|||||
Db 20528 GTGATCTGTACGTATCGCAATCTCTGTAAACAAACCAAGGCTATAAAGCGCTGCT 20469

QY 664 ttcccttgcacacagacgctggcgttagcaacagctactgtggaacaaagctcgcgaccatc 723
|||||
Db 20468 TTCCCTTGCACACAGACGCTGGCGTAGCAACAGCTACTTGGAAACAAAGTCTGCGACCATC 20409

QY 724 aattatcatgaatggcaatgagagcctctctattctacagactaaactcttagtgcca 783
|||||
Db 20408 AATTATCATGAATGGCAAGTAGAGCCTCTCTATCTTACAGACTAAACCTTTTAGTGCCA 20349

QY 784 tacattggagtaacaattgctcagcaactttttagctgataaactcgcattcgtcag 843
|||||
Db 20348 TACATTGGAGTACAATGGTCTCGAGCAACTTTTGATGCTGATACATCCGCATTTGCTCAG 20289

QY 844 ccaaaactaacactagctgttttaaaacttaactgaatgaacccctcttactaggaat 903
|||||
Db 20288 CCAAAACTACCTACAGCTGTTTTAAACTTAACTGATGATGAACCCCTCTTTACTAGGAAT 20229

QY 904 gccacagcattgctactactatgcttctcagacttcactcaaatgttctcctgtag 963
|||||
Db 20228 GCCACAGCATGTGCTACTACTGATTCGTTCTCAGACTTCATGCAAAATGTTTCCCTGTCAG 20169

QY 964 atcaacaagttaaatctagaaaaagctgtgaggttactgtgaggagctacttagttgat 1023
|||||
Db 20168 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 20109

QY 1024 gctgataaatggtcaacttactcagaaagctgtttaaattaaagagagagctcctcaagta 1083
|||||

Db 20108 GCTGATAATGGTCACTTACTGCAGAAAGCTCGTTTAAATTAACGAGAGAGCTGCTCACGTA 20049

QY 1084 tctggtcaggttcagattc 1101
|||||

Db 20048 TCTGGTCAAGTTCAAGATTC 20031

RESULT 6

AP002547 300550 bp DNA BCT 08-DEC-2000

LOCUS Chlamydomydia pneumoniae J138 genomic DNA, complete sequence, section 3/4.

ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008

VERSION AP002547.2 GI:10176693

KEYWORDS

SOURCE Chlamydomydia pneumoniae J138 (strain:J138) DNA.

ORGANISM Chlamydomydia pneumoniae J138.

REFERENCE 1 (sites)
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.

AUTHORS Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.

TITLE Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA

JOURNAL Nucleic Acids Res. 28 (12), 2311-2314 (2000)

MEDLINE 20330349

REFERENCE 2 (bases 1 to 300550)

AUTHORS Shirai, M.

TITLE Direct Submission

JOURNAL Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases. Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227, Fax:81-836-22-2415)

COMMENT On Sep 15, 2000 this sequence version replaced gi:6172298 gi:6172300 gi:6172396 gi:6172398 gi:8978889.

FEATURES

source

1. 300550

/organism="Chlamydomydia pneumoniae J138"

/strain="J138"

/db_xref="taxon:138677"

/note="synonym:Chlamydia pneumoniae (strain J138)"

gene

complement(21..794)

/gene="dapF"

CDS

complement(21..794)

/gene="dapF"

/codon_start=1

/transl_table=11

/product="glutaminolipase epimerase"

/protein_id="BAA98725.1"

/db_xref="GI:8978890"

/translation="MAFSPSTISKYIYSGAGNRFLIGETLPEVDVRFLEQTRVD GFYLPKSSCADAOLITFNSDGRPTWGNLCAIAHLASOKSKSDISYSTDGLYS GYFYSWRDLVDTLADWRASVHRLESPPDLPEKVCIHGTGPHAVVILPEISLDL SILGFLRYHTFSPDGNVNVNVLGLHCQLRVRYERGVETGACGTGALASALVV SNSYWKESIQIHTWGGELMTVSQNRGRVYLQGSVTHDL"

gene

complement(763..1338)

/gene="clpP_1"

CDS

complement(763..1338)

/gene="clpP_1"

/codon_start=1

/transl_table=11

/product="CLP protease"

/protein_id="BAA98726.1"

/db_xref="GI:8978891"

/translation="MADGEVHKLPDIIKELEARRVFEPPVTEKSASDAIKKLWYL ELKDPGRPIVFVINSFGSVNDAGFAVNDQIKMLSPVTVTYGLAASNGSVLSCLAP GRRTATPSRIMIHOPISIGGPTITGQATDLDIHAREILKTRILNDIVTEATNQPDII EKAIDRWMTANEAKDFGLDGLIFSNFL"

gene

complement(1358..2917)

/gene="glyA"

CDS

complement(1358..2917)

/gene="glyA"

```

/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/transl_table="MLKVFKEFKFAIVEIFTKVAVVSVLLHKLLENASGKKGOSLAS
TAYLAADHLNAPSPISGERIIDELKSORSHLKMIASENYSLSVOLAMGNLLTDKYC
EGSPKFRKYSCEENVDAIEWECVETAKELPAADCACVOPHSGADANLLAVNAILTHKV
OGPAVKLGKTYTVELTEEBEYTLKAESSVCVCLGPSLNSGCHLTHGNVRNUNVSKLM
RCFPYDVNTECEDYAEISRLAKEIKPKVLIAGTSSYRRRLNFAVLKQIAEDCGSVL
WYDMAFAGLVAGGVFVDEENPIYADIIVTTTTHTKLRGPRGGVLATREYESTLNKA
CPLMGGSPFHVIAAKTVALKEALSVDKFKYAHQVNNARLAEFSLHGLRLTGGT
DNHMMVIDLGSIGSGIAEDILSSVGIAVNRNSLPDAIGKWTSGIRLGTALTTG
GMGIDEMEADVADIIIVKVLNRNLSCHVEGSSKKNKGELPEAIAQEARDRVNRNLLRFP
LYPTDLEALV"
gene
3107..3778
/gene="CPj0522"
CDS
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/transl_table="MTLYLGNKTKARKYQHYLPILTLFPYAKSTPQNKRALQFLPQ
ATHVILSPSTHFLSRMTSLLSKATLTKTYLCIGESTKERILLFSGQVYVYVATQ
ETAEIIFPLQALPSSARILYPHSSLARPVTREFLYNRTFYSYPHYTVPRKLLKNI
LSKYKKIIFTSPSTVRAPAKIFPREPKTYWCQGMTLQEFQKFSQKQVSLLETICK
SRTSP"
gene
3843..4175
/gene="CPj0523"
CDS
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/transl_table="MASSATPGDGTAPSLPFPATPRPNFKALFVTAIALVWIAL
IATTAIGLCIHPLCSFIFLPAIFLYFSRYICSHYARNVYIALDVDPHSLQDMRS
HSPIFSDR"
gene
4223..5302
/gene="CPj0524"
CDS
4223..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/transl_table="MSGPSRTSSQSVLSYVPRDKETAPRKQFTIAKISTAILASL
ALGALVAGISITVILGNPVFLALITLTFVSVTVFLVYHOMTSKVSNQKVLQNKPK
PLGRWQEKYKDYCSNEMQVNNHLPKFKVAIOTDASQFPQFTLGLRVIEKNQST
GILFNPGPTNLNDNTNLSTILYSLDKSVMDTKQREGGPAKGEDPSPTEVRV
VKLPNEALQDFNLNLSAEKKSILPFLGHVCGPKSEELPNQEQYTHQALLAYENCL
KAALIESHAIVALPLFTSVIEVPPEILPKEGTFTYDNTQTAQFCRALLDAIQNTALR
YPOKSLVILQDPNTIESQSRSE"
gene
5642..6406
/gene="CPj0525"
CDS
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/transl_table="MHDALLSILAQELDKMIRLMRVKKEHOKELAKVQSKSDIRR
RVQKELEMEENLKTQIRDGENRIQISEQINKLENQQAQVKKMDEFNALTOEMTANK
RRSLHQLSDMLQKAGGEDLYSLKESLASTENSSVIEKEIFESTIKINEEGKAL
LEQRLKHAHNPELLSIYERLLNKKDRVVVPTENRVCSGCHIVLTPOHENLVRRKD
RLIFCEHCSRILYQESQVNAQENSTAKRRRAAV"
complement(5827..7816)
/gene="kpsf"

```

```

CDS
complement(6827..7816)
/gene="kpsf"
/codon_start=1
/transl_table=11
/product="GutQ/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/transl_table="MPSPMISTDVQCQILGKQKQKAVDFFOAFQPKAMQALAEILGH
SGWFFSGVGSGCVARKLVATLQSLSERALFFSPDLLHGDGLGVSPGDIIVCLFSKS
GETOELDTVPHLKSRAILVAITSMYPNLAALSDLVILPSVAELDPFNLIPTNST
TCOMIFGDFLAMELHFSRGVSLSTYGNHPSQGVGMKANGKVKDFMFPKTEVPFCHLG
DKVFSFSLGAGCGVCIVDPQFLMGIFTDGLRRSLASYGGEVLSLSLEKVMTA
NPKCITEDSDIAIALQLMESSPVAVLVPLDNEENRHVTGLLHMHITLAKACLL"
gene
complement(7843..9024)
/gene="sucB_2"
CDS
complement(7843..9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/transl_table="MIFEFPPKIGETSSGGSIVRWLKNLGDHVARDEPLIEVSTDKI
ATELSPKAGRLRVFCVNEGDEVASGDVLGLEIEEISEADESTSCPLTSCETKSEA
GSSSSVMFSPAVLSLAQREGIDLNQKIAGTGGKGRVTRQDLEAVISEQQVSIPE
IFQCEVNRIPMSPLRRRAIASSLSKSDSEVHASLVVDVDTDLNMLINSGERQRFDLTH
GVKITITFSIVQCLAOTLRQPLANGSLDGTIVMKSVNVGVAVLNKKEGVVPLVH
NCQDGLVSIKALADLSSRARLNKLDSEVQDSVTVNFCMTGALIGMFIIRYPEV
AILGIGTIQRVVVRRDDDSLAIKRMVVVTLTFDHRVLDGIYSEBFLTSLKNRLESVTM
G"
gene
complement(9038..10282)
/gene="gltr"
CDS
complement(9038..10282)
/gene="gltr"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/transl_table="MKLMKIFIGLVGVTLGLVLEDKAIEFKPIGIDIFLNLNVVY
PLVCSWAVTASISDKMKKIGIGISVGLVLTITLATAIYGLCFATYFSPNGCDFA
QAQSDMSAVTIDSNKTAAYFLSIHQVPSNPVPSFAEGNIIQIIFAILGIALRL
SGERGPVERIDGSEIMLRNMNIMSFAFYGVGASMAWISGHGSLVWLQGLKFI
AYLACLFLHATLVFGGLVFCGKMSFSLSSMDAISCAVSTASSATLPLVTWRCVS
KNLGSVAEVSFVLPLGATVNMGTALFOGAAVFAQAYNCPLSLSLLLVVTATF
SAVGSAGVPGGEMITLGSVLASVGLPIQGIALLAGIDRLDIVGTPMNLGDVAVVATY
VASGEGLSPYESIKQESVET"
gene
complement(10282..11379)
/gene="ycha"
CDS
complement(10282..11379)
/gene="ycha"
/codon_start=1
/transl_table=11
/product="Atpase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/transl_table="MKKRVSPFTLFFYRRVTAISLEIGLGMWGLSLLSKVFAFLVA
CWNRFNSTPYRARSTVIVGNVIVGGAGKTPVILWLAELRLRGYSGVLSRGYKSO
SSROKLTIVDSKVHSASVGDPELLMAEKLPESVWVHKDRRIISAARAEKFGILL
DDGLQYRLKHADVIAVYVQDPLGKRAFFPKGRURDFPLRLKTVDAIIVNGGKEAG
TVKRVSNAPQIFVKRYTASVWTHGERIPKEALRELVRGVFCGLGPGFQGLNMLRE
EGHILGLYLLPDHAAITKELNFTFCQOMANROGGLCTEKDYSKVLPRLSGEVSLP
IAKVMRLSVNQDDTSLNLMIEQIHKNRGN"

```

Query Match

99.7%; Score 1098; DB 1; Length 300550;

Best Local Similarity 100.0%; Pred. No. 1.3e-215;

Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 4 ttgcctgtaggaacccctctctgatacgaacttaataatgattggtacataatgggaagt 63

DB 179406 TTGCCTGTAGGGAACCCCTCTCTGATCCAGCTTATTAAATGATGTTACATATGGAAGT 179465

Qy 64 gctgcagagatccttgogactctgcctacttggtgcgcgacttagctacgtgct 123

```

|||||
Db 179466 GCTGAGAGAGATCTTGGCATCTTGGCTACTTGGTGGCGAGCTATTAGCTTACGTGCT 179525
Qy 124 ggattttacgagactatgttttgcacgctatctttaaagtagatgcacctaataacattt 183
Db 179526 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGACCTTAAACATTT 179585
Qy 184 tctatggagcagcactactggatccgctgctgcacaaactactactacgcgtagataga 243
Db 179586 TCTATGGAGCAAGCCACTTGGATCGCTGCTGCAAACTACTACTGCGGTAGATAGA 179645
Qy 244 cctaaccgagcctacaataaagcattacacgatgcagagtggttcactaaatgcaggcttc 303
Db 179646 CCTAACCCGGCTACAAATAGCAATTTACAGATGCAGAGTGGTTCACATAATGCAGGCTTC 179705
Qy 304 attgcttaaacatttgggagctgcttggatgtttcttctgactttaggagctcttaaggt 363
Db 179706 ATTGCTTTAAACATTTGGGATCGCTTTCATGTTTCTCTACTTTAGGAGCTTCTAATGGT 179765
Qy 364 tacattagagaaactctacagcgttcaatctcgttggttttatttcgaggttaaggtact 423
Db 179766 TACATTAGAGAACTCTACAGCGTCAATCTCGTGTGTTTATTCGGAGTTAAAGGTACT 179825
Qy 424 actgttaaatgcaaatgaaactacaaacacgtttctttaaagtaacgaggtgtttgaactttac 483
Db 179826 ACTGTAAATGCAATGAATACCAACCGTTCTTTAAGTAACGAGTGTGTTGAACCTTAC 179885
Qy 484 acagacaccttctcttgagcgttaggcgtctgagccttatgggaatgcggttct 543
Db 179886 ACAGACACCTCTTCTTGGAGCGTAGCGCTGCTGGAGCGTTATGGAATGCGGTTGT 179945
Qy 544 gcaactttgagagctgaattcacaatatacacagctcacaactaaagtgaagaacttaac 603
Db 179946 GCACTTTGGGAGCTGAATTCCAATATGCACAGTCCAACTCCAACTTAAAGTTTGAAGAACTTAAT 180005
Qy 604 gtgatctgaacgtatcgcaattctctgttaaaacacccaggggctataaaggcgttgc 663
Db 180006 GTGATCTGAACGTATCGCAATCTCTGTAAACAAACCCAGGGGCTATAAAGGCGCTTGT 180065
Qy 664 tccctctgcaacagagcgtgcgttagcaacagctactggaacaaactctgcacaccac 723
Db 180066 TTCCCTCTGCCAACAGACGCTGGCTAGCAACAGCTACTGGAACAAAGTCTGCGACCAATC 180125
Qy 724 aattatcatgaatggaagtagagcctctctctatcttacagactaaactcttttagtgcca 783
Db 180126 AATTATCATGAATGGCAAGTAGAGCCCTCTATCTTACAGACTAAACTCTTTAGTGCCA 180185
Qy 784 tacattggagtaacaatggtctgcagcaacttttgatgctgtataacatcgcattgtctcag 843
Db 180186 TACATTGGAGTACAATGGTCTCGAGCAACTTTTGTGCTGTATGATAACATCGCATTTGCTCAG 180245
Qy 844 ccaaaactacctacagctgttttaaaacttaactgcatggaacccctctttactagaaat 903
Db 180246 CCAAAACTACCTACAGCTGTTTAAACTTAACGTGATGATGGAACCCCTCTTTACTAGAAAT 180305
Qy 904 gccacagcattgtctactactgattctgtctcagactctcatgcaaatgttttcctgtcag 963
Db 180306 GCCACAGCATGTCTACTACTGATTCGTTCTCAGACTTTCATGCAAAATGTTTTCCTGTGCTCAG 180365
Qy 964 atcaacaagttaaatctctagaaaaagctgtgaggttactgtgaggtactcttagttgat 1023
Db 180366 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 180425
Qy 1024 gctgataaatggtcactactcagagcgtctgttaataataacgagagagcgtctcagcta 1083
Db 180426 GCTGATAAATGGTCACTTACTCGAGAAAGCTCGTTTAATTAACGAGAGAGCTGCTCACGTA 180485
Qy 1084 tctggtcagttcagattc 1101
Db 180486 TCTGGTCAGTTCAGATTC 180503
RESULT 7

```

```

CHTMOMP 1170 bp DNA BCT 26-APR-1993
LOCUS Chlamydia pneumoniae major outer membrane protein (MOMP) gene,
DEFINITION complete cds.
ACCESSION M64064 M34942 M64063
VERSION M64064.1 GI:144534
KEYWORDS major outer membrane protein.
SOURCE Chlamydia pneumoniae (strain IOL-207) DNA.
ORGANISM Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 1170)
AUTHORS Carter,M.W., Al-Mahdawi,S.A.H., Giles,I.G., Treharne,J.D.,
Ward,M.E. and Clarke,I.N.
TITLE Nucleotide sequence and taxonomic value of the major outer membrane
protein gene of Chlamydia pneumoniae IOL-207
JOURNAL J. Gen. Microbiol. 137, 465-475 (1991)
MEDLINE 91237311
FEATURES
Location/Qualifiers
1..1170
/organism="Chlamydia pneumoniae"
/strain="IOL-207"
/db_xref="taxon:83558"
1..69
/gene="MOMP"
1..1170
/gene="MOMP"
/codon_start=1
/transl_table=11
/product="major outer membrane protein"
/protein_id="AA23143.1"
/db_xref="GI:144535"
/translation="MKLLKSLISAAFGVSGLSQALPVGNPSDPSLLIDGTIWECA
AGDPCDCAWCDALISLRAFGYDVFDRILKVDAPKTSMGAKPTGSAANVTYAVD
RNPAYNKLHLDIAEWFNAGFIALNLDWDFVCTLGASNGYLRGNSATNVLGLFV
KQTVNANELPNVSLNGVVELYTDTSFSSVSGARGALWEGCATLGAFOYAOQSPK
VEELNVICNVQSFSVNPAPKGVAFPLPDAGVATATGKATINHEWQVGLSLY
RLNSLVPYIGVQWNRATFDADNIRIAQPLIPTAVLNLTAWNPSSLGNATSLSTDSFS
DFMQIVSCQINKFKSRKACGVTGATLVDAKWSLTAEARLINERAHVSGQFRF"
1..1170
/gene="MOMP"
70..1167
/gene="MOMP"
/product="major outer membrane protein"
BASE COUNT 319 a 255 c 247 g 349 t
ORIGIN
Query Match 99.6%; Score 1096.4; DB 1; Length 1170;
Best Local Similarity 99.9%; Pred. No. 2.8e-275;
Matches 1097; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 4 ttgcctgtaggaaccccttctgacccaagcttatttaattgattggtacaaatgggaaggt 63
Db 70 TTGCTGTAGGGAACCCCTCTGATCCAAAGCTATTAAATGTGTTACAAATATGGAGGT 129
Qy 64 gctgcagagatccttgcgactcttgcctacttggctgagcgtctattagcttaagctgct 123
Db 130 GCTGCAGGAGATCCTTGGATCCTTGCCTACTTGGTGCAGCGCTATTAGCTTACGTCT 189
Qy 124 ggattttacgagactatgttttgcacgctatctttaaagtagatgcacctaataacattt 183
Db 190 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGCGACCTTAAACATTT 249
Qy 184 tctatggagcagcactactggtatccgctgctgcacaaactactactacgcgtagataga 243
Db 250 TCTATGGAGCAAGCCACTTACGATCCGCTGCTGCAAACTACTACTTACCTAGATAGA 309
Qy 244 cctaaccgagcctacaataaagcattacacgatgcagagtggttcactaaatgcaggcttc 303
Db 310 CCTAACCCGGCTACAAATAGCAATTTACAGATGCAGAGTGGTTCACATAATGCAGCTTC 369
Qy 304 attgcttaaacatttgggagcgttctgattgtttctgactttaggagcgtcttaaggt 363

```



```

RESULT 11
AF131230      670 bp      DNA      BCT      01-JUN-1999
DEFINITION   Chlamydothila pneumoniae mutant major outer membrane protein (MOMP)
              gene, partial cds.
ACCESSION    AF131230
VERSION      AF131230.1 GI:4928269
KEYWORDS     Chlamydothila pneumoniae.
SOURCE       Chlamydothila pneumoniae.
ORGANISM     Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.
REFERENCE    1 (bases 1 to 670)
AUTHORS      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
TITLE        Presence of viable Chlamydia pneumoniae in fetal calf serum and in
              epithelial-derived cell lines
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 670)
AUTHORS      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
TITLE        Direct Submission
JOURNAL      Submitted (25-FEB-1999) Pathology, Vanderbilt University, C-3321
              Medical Center North, Nashville, TN 37232, USA
FEATURES     Location/Qualifiers
              1..670
               /organism="Chlamydothila pneumoniae"
               /db_xref="taxon:83558"
               /note="synonym: Chlamydia pneumoniae
               isolated from commercial fetal calf serum"
               <1..>670
               /gene="MOMP"
               /note="MOMP"
               <1..>670
               /gene="MOMP"
               /note="MOMP"
               /note="similar to MOMP gene sequences of Chlamydia
               pneumoniae presented in GenBank Accession Numbers M64064,
               M34942, and M64063"
               /codon_start=1
               /transl_table=11
               /product="mutant major outer membrane protein"
               /protein_id="RAD33511.1"
               /db_xref="GI:4928270"
               /translation="VDRPNPAYNKHLDHDAEFTNAGFIALNIWDRFDVFTLGNNGY
               IRGNSTAFNLVGLFVGKVTNNANELFNSLNSGVELYDTDFSWSGVARGALWECG
               CATINGEAFQAQSPKVEELNVICNVQSFYKPKYGVAFPLPTDAGVATGTGS
               ATINHEWQVGSLSYRLNSLVPYIGVQWSRATEDADNIRIAQPKLPTAVLNLTANP
               SLLGN"
               63
               /gene="MOMP"
               /note="difference from wild type, silent mutation"
               /replace="a"
BASE COUNT   191 a 147 c 138 g 194 t
ORIGIN
Query Match 60.7%; Score 668.4; DB 1; Length 670;
Best Local Similarity 99.9%; Pred. No. 1.le-163;
Matches 669; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 235 gtagatagacctaacccggcctacaataagcatttacagatgcagagtgggttcaataat 294
DB 1 GTAGATAGACCTTAACCGCGCTACAAATAGCATTTACAGATGCAGAGTGGTTCACATAAT 60
QY 295 gcaggcttcattgccttaaacatttgggagcgtttgattgtttctgtacttttagagct 354
DB 61 GCGGGCTTCATGGCTTAACATTTGGGATCGCTTTGATGTTTCTGTACTTTAGAGCT 120
QY 355 tctaattggttacattagaggaactctacagcgttcaatctcgttgggtttattcgagtt 414
DB 121 TCTAATGGTTACATTAGAGAACTCTACAGCGTTCAATCTCGTTGGTTTATTTCGGAGTT 180
QY 415 aaaggtactactgtaataacaaactacaaacgtttctttaaagtaacagagttgtt 474
DB 181 AAAGGTACTACTGTAATGCAATGAACATACCAACAGTTTCITTAAGTAACGAGTTGTT 240

```

```

QY 475 gaactttacacagacacctcttctcttgagcgtgagcgtcgtgagcgttatggaa 534
DB 241 GAACTTTACACAGACACCTCTTCTTCTTGAGCGTAGCGCTCGTGAGCCTTATGGAA 300
QY 535 tgcggtgtgcaactttggagcgtgaattccaataatgcacagtcaccaacacctaagttaa 594
DB 301 TGCGGTTGTGCACACTTTGGGAGCTGAATTCGAATATGCAGTCCAAACCTTAAAGTTGAA 360
QY 595 gaacttaattgtatcgttaacgtatcgaattctctcttaaaacacccagggcgtataaa 654
DB 361 GAACTTAATGTGATCTGTACGTATCGCAATTCCTGTAAACAAACCAAGGGCTATAAA 420
QY 655 ggcgttcttcccttgcacacagacgctggcgtagcaacagcagctactggaacaaagtct 714
DB 421 GCGCTTCTTCCCTTGGCCAAACAGACGCTGGCGTAGCAACAGCTACTGGAACAAAGTCT 480
QY 715 ggcaccatcaattatcatgaatggcaagtaggagcctctctcttaccagactaaactct 774
DB 481 GCGACCATCAATTATCATGAATGGCAAGTAGGAGCCTCTCTATCTTACAGACTAAACTCT 540
QY 775 ttagtgcatacttgagtagacaatgctcgcagcaacttttgatgctgataacatcgc 834
DB 541 TTAGTGCCATACATTTGGAGTACAATGGTCTCGAGCAACTTTTGTATGCTGATACATCCGC 600
QY 835 attgctcagcaaaactacacacagcgtgttttaaaacttaacgtcagcgaacccctcttta 894
DB 601 ATTGCTCAGCAAAACTACCTACAGCTGTTTTTAAACTTAACATGCAACCCCTTCTTTA 660
QY 895 ctaggaaatg 904
DB 661 CTAGGAAATG 670

```

RESULT 12

```

AF131229      670 bp      DNA      BCT      01-JUN-1999
LOCUS        Chlamydothila pneumoniae mutant major outer membrane protein (MOMP)
DEFINITION   gene, partial cds.
ACCESSION    AF131229
VERSION      AF131229.1 GI:4928267
KEYWORDS     Chlamydothila pneumoniae.
SOURCE       Chlamydothila pneumoniae.
ORGANISM     Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.
REFERENCE    1 (bases 1 to 670)
AUTHORS      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
TITLE        Presence of viable Chlamydia pneumoniae in fetal calf serum and in
              epithelial-derived cell lines
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 670)
AUTHORS      Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.
TITLE        Direct Submission
JOURNAL      Submitted (25-FEB-1999) Pathology, Vanderbilt University, C-3321
              Medical Center North, Nashville, TN 37232, USA
FEATURES     Location/Qualifiers
              1..670
               /organism="Chlamydothila pneumoniae"
               /db_xref="taxon:83558"
               /note="synonym: Chlamydia pneumoniae
               isolated from cryptically-infected H292 cells"
               <1..>670
               /gene="MOMP"
               /note="MOMP"
               <1..>670
               /gene="MOMP"
               /note="similar to MOMP gene sequences of Chlamydia
               pneumoniae presented in GenBank Accession Numbers M64064,
               M34942, and M64063"
               /codon_start=1
               /transl_table=11
               /product="mutant major outer membrane protein"
               /protein_id="AAD33510.1"
               /db_xref="GI:4928268"
               /translation="VDRPNAAYNMHLHDAEFTNAGFIALNIWDRFDVFTLGNNGY

```

IRNSTAFNLVGLFGVKTNNANLNPVLSLNGVVELYTDTSFWSVYGARGALWECC
CATLGAEOYAGSKPKVEALNVICNVQSFNSKPKYGVAFPLEPTDAGVATATGKS
VTIYNHGVQVGLSLVRLNVLPIGVQVSRATFDADNIRIAQPKLPVAVNLTAWNP
SLVGN"

variation

16
/gene="MOMP"
/note="difference from wild type; in wild type encodes
Pro"
/replace="c"
29

variation

/gene="MOMP"
/note="difference from wild type; in wild type encodes
Lys"
/replace="a"
482

variation

/gene="MOMP"
/note="difference from wild type; in wild type encodes
Ala"
/replace="c"

BASE COUNT 191 a 144 c 139 g 196 t
ORIGIN

Query Match 60.4%; Score 665.2; DB 1; Length 670;
Best Local Similarity 99.6%; Pred. No. 7.3e-163;
Matches 667; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 235 gtatagatagcctaaccgcccctacaataagcatttacagatgcagagtggttcactaat 294
DB 1 GTAGATAGACCTAACCGCGCCCTACAAATATGCATTTACACGATGCAGAGTGGTTCATAAT 60
QY 295 gcaggcttcattgccttaaacatttgggacgcttgcattgtttctgtacttagagct 354
DB 61 GCAGGCTTCATTGCCCTTAACATTTGGGATCGCTTGTATGTTTCTGCTACTTTAGAGCT 120
QY 355 tctaattggttacattagagaaactctacagcgttcaactcgttgggtttattcgagtt 414
DB 121 TCTAATGGTTACATTAGAGAACTCTACAGCGTTTCAATCTCGTTGTTTATTTCGGAGTT 180
QY 415 aaaggtactactgtaataacaaactacaaactacaaactcttcttaagtaacgagtggt 474
DB 181 AAAGGTACTACTGTAATAATGCAATGCAATGCAATGCAATGCAATGCAATGCAATGCAAT 240
QY 475 gaacttacacagacacctcttctcttcttcttcttcttcttcttcttcttcttcttcttct 534
DB 241 GAACCTTACACAGACACCTCTTCTCTCTGAGCGTAGGCGCTCGTGGAGCCTTATGGAA 300
QY 535 tgcggttgcaacttggagctgaattcccaatatgcacagtcacaaacccaaagtgaa 594
DB 301 TGGCGTGTGCAACTTTGGAGCTGAATTCGAATATGCACAGTCCAAACCTAAAGTTGAA 360
QY 595 gaacttaattgattgatacgtatcgcaattctctgtataacaaacccaaagtgctataa 654
DB 361 GAACCTAATGATCTGTAACGTATGCAATCTCTGTATAACAAACCAAGGCGCTATAA 420
QY 655 ggcgtgttcccttgcacacagacgctggtgtagcaacagctactggaacaaagctc 714
DB 421 GCGGTTGCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCTTTCCT 480
QY 715 gcgaccatcaattatcatgaatggcaagtaggagctctctctctctctctctctctctct 774
DB 481 GTGACCATCAATATCATGAATGGCAAGTAGGAGCCTCTCTATCTTACAGACTAAACTCT 540
QY 775 ttagtccatcatttgagtagaattgcttcgagcaacttttgatgctgataacatccgc 834
DB 541 TTAGTCCCATACATTGGAGTACAATGGTCTCGAGCAACTTTTGTATGCTGATACATCCGC 600
QY 835 attgctcagcaaaaactacacagctgttttaaaacttaactgcatggaaccccttttta 894
DB 601 ATTGCTCAGCAAAACTACACTACAGCTGTTTAAACTTAACTGATGGAACCTTCTTTA 660
QY 895 ctagggaatg 904
|||||

Db 661 CTAGGAATG 670

RESULT 13

AF269259 1425 bp DNA BCT 07-MAR-2001
Chlamydomophila psittaci VS225 major outer membrane protein (ompa)
gene, partial cds; and tRNA-Gly gene, partial sequence.

ACCESSION AF269259

VERSION AF269259.1

KEYWORDS GI:12642454

SOURCE Chlamydomophila psittaci.

ORGANISM Chlamydomophila psittaci.

REFERENCE 1 (bases 1 to 1425)

AUTHORS Bush, R.M. and Everett, K.D.

TITLE Molecular evolution of the Chlamydiaceae

JOURNAL Int. J. Syst. Evol. Microbiol. 51 (Pt 1), 203-220 (2001)

MEDLINE 21078680

PUBMED 11211261

REFERENCE 2 (bases 1 to 1425)

AUTHORS Everett, K.D.E., Hambly, W.A. and Andersen, A.A.

TITLE Direct Submission

JOURNAL Submitted (22-MAY-2000) Medical Microbiology and Parasitology,

University of Georgia, College of Veterinary Medicine, Athens, GA

30602, USA

FEATURES Location/Qualifiers

1..1425

/organism="Chlamydomophila psittaci"

/strain="VS225"

/specific_host="orange-fronted parakeet; Aratinga

canicularis"

/db_xref="taxon:83554"

/country="USA; Laredo, Texas"

/note="Isolated from swab material submitted to APHIS in

August 1991 from a bird purchased in Laredo, TX USA;

generated with the primers CTU (5' ATGAAAAAAGCTCTTGAATCGG

3') and TGLY (5' GGCTACAGCTCTACCATTTGA 3')"

<1..1167

/gene="ompA"

<1..57

/gene="ompA"

<1..1167

/gene="ompA"

/note="porin; MOMP"

/codon_start=1

/transl_table=11

/product="major outer membrane protein"

/protein_id="AAK00240.1"

/db_xref="GI:12642455"

/translation="LLKSALLFAATGSLSLQALPVGPAEPSSLIDGTWEGAGSDP

CDPCATWCDATISIRAGYGVDFVRLVDVNTKTSIGMGAAPTGSAADYKPTDPRN

DYACKHLODAEWFNTAAFLALNIDRFDICTLGASNGYFKASSAFLNVLGLVKGKT

SVADQLPENVGITQGIYEFYDTTFSSVGARGALWEGCATLGAEFQYAOENPKIEM

LNVLSPPTGVVHKPRGKGTGNSFPLPLTAGTDGATDKSATLKYHEMQVGLALSYR

LNMLVPYIGVNMSTRATFADSIIRIAQPKAAVNLNTTNPNTLLGEATLADSNKFC

FLQIASIQINKMKSKACGAVGATLIDAKWSITGEARLINERAHHNAQPRF"

58..1164

/gene="ompA"

/product="major outer membrane protein"

1395..>1425

/product="tRNA-Gly"

primer_bind 394 a 313 c 294 g 424 t

BASE COUNT complement(1406..1425)

ORIGIN

Query Match 55.1%; Score 606.6; DB 1; Length 1425;

Best Local Similarity 73.1%; Pred. No. 1.4e-147;

Matches 809; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

QY 4 ttgcctgtagggaaccccttctgatccaagcttattgaatggtggaaggt 63

|||||


```
Db 187 GCATACACGAGATATGTTTTCGATCGTGTATTAAAGTTGATGTAATAAACTATC 246
Qy 184 tc---tatggagcaagcctactgagtcgcgtgctgcaaaactatactactgcccgtat 240
Db 247 ACCGCGATGGGTGCGATGTTCTTACAGGAACCCGAGCAGCTGATATACAAACTCCTACGGAT 306
Qy 241 agacctaaacccgctacaaataagcatattacacgatgagagtggttcaactaatcagcg 300
Db 307 AGACCCACATCGCTTACGGCAACACTTACAAGAGCCGATGTTTACCAGTACGCT 366
Qy 301 ttcattgcttaaacatttgggagcgttggatgttttctgtacttttaggagcttctaatt 360
Db 367 TTTCTCGCATTAATATCTGGGATCGCTTTGATATTTCTGCACATTAGCGCTTCTAAT 426
Qy 361 gtttaccattaggaactacacgcttcacgcttcaatctcgttggtttattcggagttaaagt 420
Db 427 GGGTACTTCAAGAGCTAGTTCTCGGCAATCAACCTCGTGGTGTGATGGTGTAAAGGA 486
Qy 421 actactgtaaatgcaaatgaaactaccaaagcgtttctttaaagtaacgaggttggtaactt 480
Db 487 TCCTCGTAGCAGCTGATCAGCTTCCCAATGTAGGCATCACTCAAGGAATCGTTGAATTT 546
Qy 481 tacacagacacctcttctcttgagcgttaggcgtctgtggagccttatgggaatcgcgt 540
Db 547 TATACAGATACAACTTCTTGGAGTGTAGGTGCACGCGAGCTTTATGGAGTGTGT 606
Qy 541 tttgcaactttggagctgaattccaatatgcacagtcacacgctcaaaccttaagttgaagaactt 600
Db 607 TGTGCAACTTAAAGAGCAGAGTCCAACTAGCTGATCTTAATCCCTAAATTAATTAATGTTG 666
Qy 601 aatgtgatctgtaacgtatcgcaattctctgttaaaacaaacccaaagggctataaagcgctt 660
Db 667 AATGTAGTCTCCAGCCAGCACAATTTGTGGTTTCAAGCCCTAGAGGATACAAAGGAACA 726
Qy 661 gtttccctctgcaacagacgctgctgtagcaacagctactgtggaacaaagctcgcgacc 720
Db 727 GCATTTCTTTACCTCTACAGCTGGTACTGATCAGGCAACTGACACTAAGTTCGGCTACA 786
Qy 721 atcaattatcatgaatggcagtaggagcctctctctattcttaccagactaaactctttagt 780
Db 787 ATTAATAACACGAATGCAAGTGTGTTAGGGCTCTCTTATCGATTGAACATGCTTGT 846
Qy 781 ccatacattggagtgcaatggtctcgagcaacttttgatgctgataacatccgctattgct 840
Db 847 CTTTACATTTGGGCTAACTGGTTCAGAGCAACTTTTGTGCTGACGCTATCCGCACTGCT 906
Qy 841 cagccaaactacacagctgttttaaaacttaactgaatgcaaccccttcttactagga 900
Db 907 CAACCTAAATTAGCTGCTGCTGTGTTAACTTTGACACATGGAACCCCAACCCCTTTTAGGG 966
Qy 901 aatgccaacagcattgtctactactgattcgtttctcagacttcattgcaaatgtttcctgt 960
Db 967 GAAGCTACAACTTTAGATACCTAACCAACAAATTCGCTGACTTCTGCAATTCGCTCGATT 1026
Qy 961 cagatcaacaagtttaatacttagaagaagctgtgaggtactgttaggagctactttagt 1020
Db 1027 CAGATCAACAAATGAAGTCTAGAAAAGCTTGTGGTGTAGCTGTTGGTGTCAACGTTAATC 1086
Qy 1021 gatgctgataaagtctactactgacagaagctctgttaataacagagagagagctctcac 1080
Db 1087 GAGCTGACAAATGCTCAATCACTGGTGAAGCAGCCTTAATCAATGAAGAGCGGCTCAC 1146
Qy 1081 gtatcgtgctaggttcagattc 1101
Db 1147 ATGAATGCTCAATTCAGATTC 1167
```

RESULT 15
CHTOMPAAD
LOCUS
DEFINITION

CHTOMPAAD 1261 bp DNA BCT 07-MAY-1999
Chlamydomophila abortus strain B577 major outer membrane protein precursor (ompA) gene, complete cds.

```
ACCESSION M73036
VERSION M73036.1 GI:144564
KEYWORDS
SOURCE Chlamydomophila abortus
ORGANISM Chlamydomophila abortus
REFERENCE 1 (bases 1 to 1261)
AUTHORS Kaitenboeck B., Kousoulas K.G. and Storz J.
TITLE Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species
JOURNAL J. Bacteriol. 175, 487-502 (1993)
MEDLINE 93123168
FEATURES
    Location/Qualifiers
        1..1261
            /organism="Chlamydomophila abortus"
            /strain="B577"
            /specific_host="Ovis aries"
            /db_xref="taxon:83555"
            /note="submitted as Chlamydia psittaci"
            80..1246
                /gene="ompA"
            80..1249
                /codon_start=1
                /transl_table=11
                /product="major outer membrane protein precursor"
                /protein_id="AAD29102.1"
                /db_xref="GI:4757612"
                /translation="MKLLKSLALLFAATGTSALSLSLQALPVGNPAPSPSLIDTMMEGAS
                GDCPCSTWCDAISIRAGYGDYDFRVLKVDVKNKTITGMGAVPTGTAAAYKTPTD
                RPNYATGKHLQDAEMFTNAFALINIDRFIDFCTLGASNGYFKASAAFNVLGIVG
                KGSSIAADQLPNVGTIGTIVFTDTTDFSMVGARGALWEGCCATLGAEFYAOQSNPK
                IEMLVYSSPAQFVVVHKPRGKGTAFPLPTAGTDOATDTSATIKYHEMVGGLALSY
                RLNLVPYISVNSWRTFDADAIKIAOPKLAANLNTWNPNTLLGATALDTSNKPA
                DFLOIASIQINKMKSRKACGVAVGATLIDADKNSITGEARLINERAHHMAHQFRF"
BASE COUNT 359 a 275 c 277 g 350 t
ORIGIN
    Query Match 54.5%; Score 600.2; DB 1; Length 1261;
    Best Local Similarity 72.2%; Pred No. 6.6e-146;
    Matches 795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;
    Qy 4 ttgctgttagggaaccccttctgataccaaagcttatttaattgattggtgacaaatgggaaggt 63
    Db 146 TTGCTGTAGGGAACCCAGCTGAACCAAGTTTATTATCGATGGCAGCTATGTGGGAAGT 205
    Qy 64 gctgcagagatccttgcgactcttgcctacttgcgtgcgcgcgtattagcttagctgct 123
    Db 206 GCTTCAGGTGATCCTTGGGATCCTTGTCTACTTGGTGTGATGCTATCAGCATCGCGCA 265
    Qy 124 ggattttacgagactatgttttcgacgctatctttaaagtagatgcacctaataacattt 183
    Db 266 GGATCTACTAGGAGATATGTTTTCGATCGTGTATTAAAGTTGATGTAATAAACTATC 325
    Qy 184 tc---tatggagcaagcctactgagtcgcgtgctgcaaaactatactactcgcgtatag 240
    Db 326 ACCGCGATGGGTGCGATGTTCTTACAGGAACCCGAGCAGCTAATTAACAACTCCTACGGAT 385
    Qy 241 agacctaaacccgctacaaataagcattacacgatgagagtggttcaactaatcagcg 300
    Db 386 AGACCCACATCGCTTACGGCAACACTTACAAGAGCCGATGTTTACCAGTACGCT 445
    Qy 301 ttcattgcttaaacatttgggagcgttggatgttttctgtacttttaggagcttctaatt 360
    Db 446 TTTCTCGCATTAATATCTGGGATCGCTTTGATATTTCTGCACATTAGCGCTTCTAAT 505
    Qy 361 ggttaccattaggaactacacgcttcacagcttcaatctcgttggtttattcggagttaaagt 420
    Db 506 GGGTACTTCAAGAGCTAGTTCTCGGCAATCAACCTCGTGGTGTGATGGTGTAAAGGA 565
    Qy 421 actactgtaaatgcaaatgaaactaccaaagcgtttctttaaagtaacgaggttggtaactt 480
```

Db 566 TCCTCCATAGCAGCTGATCAGCTCCCAATGTAGGCATCACTCAAGGAATCGTTGAATTT 625
 Qy 481 tacacagacacacctcttctcttgagcgttaggcgctctgagccttatgggaatgcggt 540
 Db 626 TATACAGATACAACATCTCTTGGAGTGTAGGTGCACGCGAGCTTTATGGGAGTGTGGT 685
 Qy 541 tgtgcaactttgggagctgaattccaatgtcacagtcaccaacccctaaagtgaagaactt 600
 Db 686 TGTGGGACTTTAGGAGCAGAGTCCAAATACGCTCAGTCTAATCTTAAATTTGAATGTTG 745
 Qy 601 aatgtgatctgttaacgtatcgcaattctctgttaacacaccccaaggcgtataaaggcgtt 660
 Db 746 AATGTAGTCTCCAGGCCAGCACAAATTTGTGTTTACAAAGCCTAGAGGATACAAGGAACA 805
 Qy 661 gctttccctctgcacacagcgtgcgttagcagacagctactggaacaaaagtctgcgacc 720
 Db 806 GCATTTCTCTTACCTCTTAACAGCTGGTACTGATCAGGCAACTGACACTAAGTCGGCTACA 865
 Qy 721 atcaattatcatgaatggcaagtaggcctctctctctctctctctctctctctctctctct 780
 Db 866 ATTAATACCAGATGCGAAGTTGGTTAGGCTCTCTTATCGATTGAACATGCTTGT 925
 Qy 781 ccatactggagtagacaatggtctgcgagcaacttttgcgtgataacacatcgcaattgct 840
 Db 926 CCTTACATTAGCGTAAAGTTCACGAGCAACTTTTGTGATGCTGACGCTATCGCATCGCT 985
 Qy 841 cagccaaaactacacagctgttttaaaccttaactgaactggaacccctcttactagga 900
 Db 986 CAACCTAAATTAGCTGCTGCTGTGTTAACTTGACACATGGACCCCAACCCCTTTTAGGA 1045
 Qy 901 aatgcccacagcattgtctactactgattcgttctcagacttcacgaaattgttccctgt 960
 Db 1046 GAAGCTACAGCTTTAGATACTAGCAACAATTCGCTGACTTCTGCAATTTGCTTCGATT 1105
 Qy 961 cagatcaacaagttaaatactagaaaagcttgcgtgagctactgtaggagctactttagtt 1020
 Db 1106 CAGATCAACAAAATGAAGTCTAGAAAAGCTTGTGGTGTAGCTGTGGTGTCAACGTTAATC 1165
 Qy 1021 gatcgtgataaagtctactactcagaagctgtttaaataacagagagcgtctcac 1080
 Db 1166 GACGCTGACAAATGCTCATCTGCTGTGAAGCACCCTTAATCAATGAAGAGCCGCTCAC 1225
 Qy 1081 gtatctggtcagttcagattc 1101
 Db 1226 ATGAATGCTCAATTCAGATT 1246

Search completed: February 7, 2002, 19:39:47
 Job time: 13901 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 20:17:12 ; Search time 685.35 Seconds
(without alignments)
1377.275 Million cell updates/sec

Title: US-09-391-606-14

Perfect score: 1101

Sequence: 1 atgttcctgtagggaaccc.....tatctgtcagttcagattc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N.Geneseq_1101.*

1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	600.2	54.5	1261	20 AAX25047	Chlamydia psittaci
2	526.4	47.8	1660	20 AAX25048	Chlamydia psittaci
3	523.2	47.5	1209	20 AAX25046	Chlamydia psittaci
4	514.6	46.7	1578	21 AAX64764	C. pneumoniae sero
5	514.6	46.7	1578	22 AAF56267	Chlamydia trachoma
6	473.4	43.5	3133	19 AAF62447	Chlamydia trachoma
7	479.4	43.5	3133	21 AAX292753	DNA encoding Chlam
8	477.8	43.4	3133	19 AAV40646	Chlamydia trachoma
9	474.6	43.1	3133	7 AAN60007	Sequence encoding
10	397.2	36.1	1038602	20 AAX201425	Complete genome se
11	352.2	32.0	744	20 AAX25045	Chlamydia psittaci

12 328.8 29.9 1047 21 AAA08124 C. trachomatis MOM
13 321.8 29.2 726 20 AAX25044 Chlamydia psittaci
14 287.4 26.1 1362 21 AAA08122 C. trachomatis MOM
15 287.4 26.1 1371 21 AAA08120 C. trachomatis MOM
16 287.4 26.1 1542 21 AAA08123 C. trachomatis MOM
17 254 23.1 1452 21 AAA08121 C. trachomatis MOM
18 205.2 18.6 831 21 AAA08125 C. trachomatis MOM
19 140 12.7 720 7 AAN60006 Sequence encoding
20 140 12.7 720 19 AAV62446 Vector lambda gtl1
21 140 12.7 720 19 AAV40645 Chlamydia trachoma
22 140 12.7 720 21 AAX292752 DNA encoding Chlam
23 87.6 8.0 936 22 AAF58252 Oligonucleotide D1
24 87.6 8.0 936 22 AAF58254 Oligonucleotide D1
25 87.6 8.0 936 22 AAF58257 Oligonucleotide D1
26 87.6 8.0 936 22 AAF58259 Oligonucleotide D2
27 87.6 8.0 936 22 AAF58262 Oligonucleotide D2
28 87.6 8.0 936 22 AAF58255 Oligonucleotide D1
29 87.2 7.9 936 22 AAF58252 Oligonucleotide D1
30 87.2 7.9 936 22 AAF58254 Oligonucleotide D1
31 87.2 7.9 936 22 AAF58257 Oligonucleotide D1
32 87.2 7.9 936 22 AAF58259 Oligonucleotide D2
33 87.2 7.9 936 22 AAF58262 Oligonucleotide D2
34 87.2 7.9 938 22 AAF58255 Oligonucleotide D1
35 75 6.8 259 15 AAQ54801 Chlamydia OMP frag
36 71.6 6.5 259 15 AAQ54800 Chlamydia OMP frag
37 54.6 5.0 245 15 AAQ54798 Chlamydia OMP frag
38 54.6 5.0 246 15 AAQ54799 Chlamydia OMP frag
39 47 4.3 50 20 AAV99685 Chlamydia OMP ge
40 47 4.3 50 20 AAV99686 Chlamydia OMP ge
41 45.4 4.1 50 20 AAV99688 Chlamydia OMP ge
42 43 3.9 75 19 AAV56996 C. trachomatis maj
43 43 3.9 75 20 AAX61262 Major outer membra
44 43 3.9 75 21 AAA08129 MOMP containing fu
45 41.2 3.7 80 14 AAQ43617 Chlamydia trachoma

ALIGNMENTS

RESULT 1
AAX25047
ID AAX25047 standard; DNA; 1261 BP.
XX
AC AAX25047;
XX
DT 05-JUL-1999 (first entry)
XX
DE Chlamydia psittaci major outer membrane protein DNA.
XX
KW Major outer membrane protein; MOMP; psittacosis; infection;
KW vaccine; genetic immunisation; ss.
XX
OS Chlamydia psittaci.
XX
FH Key Location/Qualifiers
FT CDS 80..1249
FT /*tag= a
XX
PN W09910005-A1.
XX
PD 04-MAR-1999.
XX
PF 28-AUG-1998; 98WO-US17943.
XX
PR 28-AUG-1997; 97US-0057147.
XX
PA (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
XX
PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
XX WPI: 1999-254214/21.
DR P-PSDB: AAW98188.
XX

PT A new vaccine for Chlamydia psittaci infections
XX
PS Disclosure: Page 57-60; 72pp; English.

This DNA sequence codes for the major outer membrane protein (MOMP, see AA098187) of Chlamydia psittaci strain B577. A claimed MOMP polypeptide (see AA098184) comprises regions VD3 and VD4 of MOMP, i.e. it lacks regions VD1 and VD2. A claimed vaccine composition includes the MOMP polypeptide, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing C. psittaci infection by administering the vaccine containing the MOMP polypeptide. Vectors encoding MOMP polypeptides lacking regions VD1 and VD2 are useful for genetic vaccination. The vaccines are used to prevent C. psittaci infection, especially in birds.

Sequence 1261 BP: 359 A: 275 C: 277 G: 350 T: 0 other;

Query Match	54.5%	Score 600.2;	DB 20;	Length 1261;
Best Local Similarity	72.2%	Pred. No. 6.le-173;		
Matches 795; Conservative	0;	Mismatches 303;	Indels 3;	Gaps 1;

Qy	4	ttgctgtagggaaccccttctgtatccaagcttatttaattgatggtacaatatagggaaggt	63
Db	146	ttgctgtagggaacccagctgaaccaagtttattaatcgtatggcactatgtgggaaggt	205
Qy	64	gctgcaggagatccttgcatccttcgcgtacttggtcgcgcgtattagcttagctgtct	123
Db	206	gcttcagggtgatccttgcgactcttgcctacttgggtgtgtgtatcagcatccgcgcga	265
Qy	124	ggatttcacggagactatgttttcgacccgtatctctaaaaagtagtagcaoctaaaaacattt	183
Db	266	ggatactacggagattatgttttcgatcgtgtataaaaagtgatgtgaataaaactatc	325
Qy	184	tc-----tatggggccaaagcctatgataccgctgtctgcaaacatatctactcccgtagat	240
Db	326	accggcattgggtgcagttcctcaggaacccgcgcgcagctaattacaaaaactctacggat	385
Qy	241	agacctaacccgcctacaataagcatttacacgatgcgcagatgggttcactaatgcagc	300
Db	386	agaccacaatcgccttacggccaacacttacagagcgcgcgaatgggttcaccaatgcagct	445
Qy	301	ttcattgccttaaacatttgggacgccttgatgtttcttctgtacttttaggagcttcta	360
Db	446	ttctcgattgaatactatcgggacgccttgatattttctgcacattaggcgcttcta	505
Qy	361	ggttacattagaggaaactctcacagcttcaatctcgttggtttatctcgaggttaaaggt	420
Db	506	gggtacttcaaaagttagtcttgggcattcaacctcgttgggttgggttgaagga	565
Qy	421	actactgtaaatgcaaatgaactacccaacggtttctttaaagtaacggagtttgttgaactt	480
Db	566	tcctccatagcagctgacagctcccaatgttaggcatacactcaaggaaatcgttgaatt	625
Qy	481	tacacagacacctttctcttgagcgttaggcgctcgttggagccctatgggaatcggt	540
Db	626	tatacagatacaaacattctcttggagtgatagggtgcacgcggagctttatgggagtggtt	685
Qy	541	tgtgcacctttgggagctgaattccaatatgcacagtcctcaaacctaaaagtgtgaagaactt	600
Db	686	tgtgcagcttttaggcagcagagtccaatcgcgcagcttaactcctaaaattgaaatgtgtg	745
Qy	601	aatgtgatctgtaacgtatcgcaattctctgtaaaacaaaccccaagggtctataaaggcgtt	660
Db	746	aatgtgtctccagcccgacacaatttgggttcacagcctagagaggtacaaaggggaa	805
Qy	661	gctttcccttgcacaagacgctggcgtagcaacagctactaggacaagctcgacc	720
Db	806	gcatttcccttaccctctaaacagctgggtacttgatcaaggcaactgacactaagtcggtaca	865
Qy	721	atcaattatcatgaatggcaagttaggagcctctctatcttacagactaaaactcttagtg	780

Db	866	attaaataccacgaaatggaagtgtggttttagcgctctctttatcgattgaacatgcttggt	925
Qy	781	ccatactggagtgacaatggtctcgagcaacttttgatgctgataaacatccgattgct	840
Db	926	ccctacattagcgtaaactggctcagcgacaactttgatgctgacgctatccgcattcgct	985
Qy	841	cagcccaaaactaactacagctgttttaaacttaactgaatgcatggaaccccttttactagg	900
Db	986	caacctaaattagctgctgtgtttaaacttgacacatggaacccacccctttttagga	1045
Qy	901	aatgccaacagcattgtctactactgattcggtctcagacttcatgcaaatgtttctcgt	960
Db	1046	gaagctacagctttagatactagcaacaataatcgctgactctctgcaaatgtgtcgatt	1105
Qy	961	cagatcaacaagttttaaattcgaaagactgtgaggttactgttaggagctactttagtt	1020
Db	1106	cagatcaacaaaatgaagctagtgaaagctgttggttagctgtttggtgcacgttaac	1165
Qy	1021	gatctgataaaatggctacacttactcagagctcgttttaattcaacgagagagctgcac	1080
pb	1166	gacgtgtgacaaattggttcataactcgtggtgaagcacgcttaataatgaaagagcgctc	1225
Qy	1081	gtatctgggtcagttccagattc	1101
pb	1226	ataaatactcaattccagattc	1246

RESULT

XX	AA25048;	
XX	05-JUL-1999 (first entry)	
XX	Chlamydia psittaci major outer membrane protein DNA.	
XX	Major outer membrane protein; MOMP; psittacosis; infection;	
XX	vaccine; genetic immunisation; ss.	
XX	Chlamydia psittaci.	
XX	Key	Location/Qualifiers
XX	CDS	364..1572
XX		/*tag= a
XX	W09910005-Al.	
XX	04-MAR-1999.	
XX	28-AUG-1998;	98WO-US17943.
XX	28-AUG-1997;	97US-0057147.
XX	(LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.	
XX	Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;	
XX	WPI: 1999-254214/21.	
XX	P-PSDB; AAW98189.	
XX	A new vaccine for Chlamydia psittaci infections	
XX	Disclosure; Page 62-65; 72pp; English.	
XX	This DNA sequence codes for the major outer membrane protei	
XX	see AAW98187) of Chlamydia psittaci strain 68C. Claimed mo	
XX	polypeptides (see AAW98183 and AAW98184) comprise regions V	
XX	of an MOMP, i.e. they lack regions VDI and VD2. Claimed va	
XX	compositions include such MOMP polypeptides, optionally fus	
XX	maltose binding protein. Also claimed are isolated nucleic	
XX	encoding the polypeptide, a vector, and a method of preventi	

CC psittaci infection by administering the vaccine containing the
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking
CC regions VD1 and VD2 are useful for genetic vaccination. The
CC vaccines are used to prevent C. psittaci infection, especially in
CC birds.
XX
SQ Sequence 1660 BP; 492 A; 342 C; 338 G; 488 T; 0 other;

Query Match 47.8%; Score 526.4; DB 20; Length 1660;
Best Local Similarity 69.5%; Pred. No. 2.3e-150;
Matches 792; Conservative 0; Mismatches 306; Indels 42; Gaps 4;

QY 4 ttgctgttaggaaccccttctgacccaagcttatttaattgattggttacaaatggaaggt 63
DB 430 ttgctgttaggaacccagctgaacaaagttatttaattgattggtgcaactgtggaaggt 489

QY 64 gctgagagagatccttgagactctgctgcctactgtgctgagcagctattgattgactgtgt 123
DB 490 gcttcaggagatccttgagactctgctgcctactgtgctgagcagctattgattgactgtgt 549

QY 124 ggaattttacgagactatgttttcacgctatcttaaaagtagatgacacataaaacattt 183
DB 550 ggaatacagagattatgttttcacgctatcttaaaagtagatgacacataaaacattt 609

QY 184 tctatgggagcgaactgactggatccgctgctgcctgcaactatctactact----- 231
DB 610 agcggcagctgctgcaactctctacgagctacaggttaacgcaagtaataactaacgcca 669

QY 232 ---gcgtagatagacctaaccgcccctacataagcatttacacgagcagagtggttc 288
DB 670 gaagcaaatggcagacccaacatcttacggaagcgcatacgaagatgagagtggttt 729

QY 289 actaatgcaggtctcattgcttaaacatttggatcgcttggatgtttctgtacttta 348
DB 730 tcaaatgcagctctcagcttaaacatttggatcgcttggatgtttctgtacttta 789

QY 349 gagcttcttaattggttacattagaggaactctacagcgttccaattcgttgggtttatc 408
DB 790 ggggcatccaattggatctactcaaacgaagtctgctgcaactcaacttgggtggttaata 849

QY 409 gaggttaaggt-----actactgtaaatgaaatgaactacacaaacgtttct 456
DB 850 ggggttctcagctgcgaagctcaactctcactgactctccaatgcaactctcactaacgtaggc 909

QY 457 ttaagtcaacggaggtgttgaaactttacacagacacactcttcttcttgagcgtaggcgt 516
DB 910 attaccgaaggtgtgtggaattttatcacagacacacatcttcttcttgagcgtaggcgt 969

QY 517 cgtgagaccttatgggaatgcggtgtgcaactttggagcgtgaattccaatatacgaag 576
DB 970 cgtgagcgtttatgggaatgtgtgtgcaactttaggagcgtgagttccaatacagctcaa 1029

QY 577 tccaacacctaaagtgaagaacttaattgctgctgtaacgtatcgcaactctcgttaaac 636
DB 1030 tctaactcctaagattgaaatgctcaacgctcaactcaagcccagcacaatttggattcaac 1089

QY 637 aaacccaagggtgtataaaggcgtt-----gcttcccttgcacacagcgtgcgcta 690
DB 1090 aaacccaagggtgtataaaggcgtt-----gcttcccttgcacacagcgtgcgcta 1149

QY 691 gcaacagcactggaacaaagctcgcacacataattatcgaatggaagcagtagagcc 750
DB 1150 acagaagctacagacacacaaacacagctcaacttaatacctgaatggcaagtagccttc 1209

QY 751 tctctattctacagactaaactctttagtgccatacttggagtgacaattgctcagca 810
DB 1210 gccctgtctacagattgaatgctgttccatatttggcgttaaacgttgcaagagca 1269

QY 811 acttttgatgctgataacacccgacttgcgcagcccaaacactacacagctgttttaaac 870
DB 1270 acttttgatgctgataacacccgacttgcgcagcccaaacactacacagctgttttaaac 1329

QY 871 ttaactgctgagaaaccccttcttactaggaaaatgccagacattgtctactact----- 924
DB 1330 attactacatggaaccccaagccttattaggatcaaacactgctttgcccaataatagtggt 1389

QY 925 ---gattcgttctcagacttcaatgcaaatgtttcctctgcagatcaacaaagtttaaatct 981
DB 1390 aaggatgttctatctgctgcttgcgaattgtctcgtgattcagatcaacaaaatgaagtct 1449

QY 982 acaaaagcttggaggttactgtagcactactttagttgattgctgataaaatggcactt 1041
DB 1450 acaaaagcttggaggttactgtagcactactttagttgattgctgataaaatggcactt 1509

QY 1042 actgcagagcgtcgttttaattaaacagagagcgtcctcagctatcgttcagttcagattc 1101
DB 1510 actggtgaagcagcgttcaatcaatgaagagcgtcctcagctatcgttcagttcagattc 1569

RESULT 3
AA25046
ID AAX25046 standard; DNA; 1209 BP.

XX AC AAX25046;

XX XX

PT 05-JUL-1999 (first entry)

XX Chlamydia psittaci major outer membrane protein DNA.

DE XX

KW Major outer membrane protein; MOMP; psittacosis; infection;

KW vaccine; genetic immunisation; ss.

XX Chlamydia psittaci.

OS WO9910005-A1.

PN 04-MAR-1999.

PD 28-AUG-1998; 98WO-US17943.

PF 28-AUG-1997; 97US-0057147.

PR (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.

PA Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;

PI WPI; 1999-254214/21.

DR P-PSDB; AAW98187.

XX A new vaccine for Chlamydia psittaci infections

PS Disclosure; Page 53-55; 72pp; English.

XX This DNA sequence codes for the major outer membrane protein (MOMP, see AAW98187) of Chlamydia psittaci strain LSUWTK, a cockatiel isolate (the MOMP gene sequence of this isolate is identical to that of C. psittaci Avian Type C). A claimed MOMP polypeptide (see AAW98183) comprises regions VD3 and VD4 of MOMP, i.e. it lacks regions VD1 and VD2. A claimed vaccine composition includes MOMP polypeptide lacking VD1 and VD2, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing C. psittaci infection by administering the vaccine containing the MOMP polypeptide. Vectors encoding MOMP polypeptides lacking regions VD1 and VD2 are useful for genetic vaccination. The CC vaccines are used to prevent C. psittaci infection, especially in CC birds.

XX SQ Sequence 1209 BP; 352 A; 263 C; 260 G; 334 T; 0 other;

Query Match 47.5%; Score 523.2; DB 20; Length 1209;
Best Local Similarity 69.3%; Pred. No. 1.9e-149;
Matches 790; Conservative 0; Mismatches 308; Indels 42; Gaps 4;

Qy	4	ttcctctgagggaacctctctgataccaaagcttataataattgatggtacaaatatgggaaggt	63
Db	67	ttgcctgtagggaacctgaaaccaagttattaaatcgatggcactatgtgggaaggt	126
Qy	64	gctgcaggagactcttgatcctctgcctactcttggtgcagcgtattagcttaacgtct	123
Db	127	gctcagagagactcttgatcctctgcctactcttggtgcagcgtattagcttaacgtct	186
Qy	124	ggatttaccgagagactatgttttcgacgctatctataaagttagatcacctaaacattt	183
Db	187	ggatactacggagattatgttttcgacgctatctataaagttagatggaataaaactttt	246
Qy	184	tctatggagccaagcctactggatccgctgctgccaactatactact-----	231
Db	247	agcggcatggctgcaactctacgcaggctaacggtaacgcaagtataactataatcagcca	306
Qy	232	---gcgctagatagaaactaaacgcgcctacaataagaatttacaacgatcagagtaggttc	288
Db	307	gaagcaaatggcagacccaactcgtctacggaagcgtatgggaagatgcagagtaggtt	366
Qy	289	actaaatgcagctctatgctttaaactattggatgcgtttgatgtttctctactttta	348
Db	367	tcaaatgcagcctctcagccttaaactattggatgcgtttgcacatttactgcacctta	426
Qy	349	ggagcttctaattggttacattagagaaactctacagcgttccaatctcgttggtttattc	408
Db	427	ggggcatccaattggatactccaagcaagttcggctgattccaacttggttgggttaata	486
Qy	409	ggagttaaaggt-----actactgtaaatgcaaatgaactacaaacagtttct	456
Db	487	gggttttcagctgaagctcaactctctaccgctattccaagcaactctcctaacgtaggc	546
Qy	457	ttaagtaacggagttgtgaacttttacagacacactcttctcttgagcgttagcgct	516
Db	547	attaccaagggtgttggaattttatacagacacatcatttctctggagcgttaggtgca	606
Qy	517	cgtggagccttatggaaatgcggtgtgcaactttggagcgtgaattccaatatgcacag	576
Db	607	cgtggagccttatggaaatgggtgtgcaacttttagagcgtgagttccaatacagctcaa	666
Qy	577	tccaacacctaaagttaagaacttaattgtatctgtaacgtatcgcaattctctgttaaacc	636
Db	667	tctaactcctaagattgaattgtcaacgtctcaactccaagccagcacaatttggattcac	726
Qy	637	aaaccacaaggcttataaaggctt-----gctttcccttggcaacagacgctggcgta	690
Db	727	aaaccagaaggcttataaaggcgtagctcgaattttctcttacctataacggttggaaaca	786
Qy	691	gcaacagctactggaacaaagctcgcagccataattatcatgaattggcaagtaggagcc	750
Db	787	acagaagctacagacaccaaatcagctacaattaaataccatgaatggcaagtaggcctc	846
Qy	751	tctctattcaagactaaactcttttagtgcatacatattggagtaacaattggctcagacca	810
Db	847	gcccgtcttacagattggaattgtgttccaatatatggcgtaaactggtccaagagca	906
Qy	811	actttgatcgtataaacatccgacttgcagcacaacactacctaacagctgttttaaac	870
Db	907	actttgatcgtataactaccgacttgcagcacaacactacctaacagctgttttaaac	966
Qy	871	ttaactgcatggaaacctcttactaggaatgcacagcattgtctactact-----	924
Db	967	attactacatggaaaccaagccttataggataccaacctgtttggccaataatagtgtt	1026
Qy	925	---gattcgtctcagaacttcgatgaaattgttctctgcagatacaacaagtttaaatct	981
Db	1027	aaggatgtctatctgatgtcttgcaattggttcgatccagatcaacaaaatgaagtct	1086
Qy	982	agaaaacctgtgaggttactgtaggagctacttttagttgatcgtataaattggtcactt	1041
Db	1087	agaaaacctgtgtgtgtactgtgtgtgtgcaacgtttaaactgcagcgtgcacaattgtcact	1146
Qy	1042	actgcagaagcctgttttaataacagagagagcgtcactcgttatctgttcagttcagattc	1101

Db |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| |
1147 actggggaagcagcttaaatcaatgaaagagtgtcacgaatgctcaattcgatttc 1206

RESULT 4
AAAG4764
ID AAA64764 standard; DNA; 1578 BP.
XX
AC AAA64764;
XX
DT 02-FEB-2001 (first entry)
XX
DE C. pneumoniae serovar MOMPS pmp gene Ral2 fusion coding sequence.
XX
KW Chlamydial infection; sexually transmitted disease;
KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
KW trachoma; blindness; acute respiratory tract infection;
KW atherosclerosis; coronary heart disease; antibacterial; ss.
XX
OS Chlamydia pneumoniae.
XX
PN WO200034483-A2.
XX
FD 15-JUN-2000.
XX
PF 08-DEC-1999; 99WO-US29012.
XX
PR 08-DEC-1998; 98US-0208277.
PR 08-APR-1999; 99US-0288594.
PR 01-OCT-1999; 99US-0410568.
PR 22-OCT-1999; 99US-0426571.
XX
PA (CORI-) CORIXA CORP.
XX
PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
XX
DR WPT; 2000-431303/37.
XX
PT Isolated polypeptide for diagnosis and treatment of Chlamydia infection
PT comprises immunogenic portion of chlamydia antigen, which comprises
PT amino acid sequence encoded by polynucleotide sequence -
XX
PS Claim 1; Page 205; 256pp; English.
XX
CC The present invention relates to new nucleic acid sequences and the
CC proteins encoded by the nucleic acid sequences. The encoded proteins
CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
CC proteins are useful for the serodiagnosis and treatment of Chlamydia
CC infection. Chlamydiae are intracellular bacterial pathogens that are
CC responsible for a wide variety of human infections. C. trachomatis
CC infection is one of the most common sexually transmitted diseases and
CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruct
CC and infertility. Trachoma due to ocular-infection with C. trachomatis
CC the leading cause of preventable blindness worldwide. C. pneumoniae i
CC major cause of acute respiratory tract infections in humans and is als
CC thought to play a role in the pathogenesis of atherosclerosis and
CC coronary heart disease. The present sequence is a nucleic acid sequenc
CC isolated in the present invention.

SQ Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;

Query Match 46.7%; Score 514.6; DB 21; Length 1578;
Best Local Similarity 68.8%; Pred.No.9e-147;
Matches 789; Conservative 0; Mismatches 314; Indels 42; Gaps

Qy 3 gttgccttgagggaaccttctgatccaagcttatcaattgatgtgtacataatgggaagg 62
|||||
Dbb 435 gctccctgtgggaatccactgactgaaccaagttttaactgatgcccatgtgsgaagg 494

Qy 63 tgcctgcaggagatccttgtgatcctctgcctactctggtagcagcatttagcttacgtgc 122
|||||
Db 495 tgcttcaggagatcctgtgcactctgcctactctgcctactctggtagcagcatttagctaccgcgc 554

QY 123 tggattttacgagactgtttttcgcgcgtatctttaaagtagatgcacctaataacatt 182
 DB 555 aggatactacgagattatgttttcgactgtttattaaagattgacgtgaataaaacttt 614
 QY 183 ttctatggagcgaacgactactggtatccgctgcaaaactataactact----- 231
 DB 615 tagcgcatggtgcgaactcctacgcaggtctatagtaacgcaagtaataactacagcc 674
 QY 232 ----gcgtatagatagacttaacccgctcctacaataagaatttacagatgcagagtgtt 287
 DB 675 aagaacaaatgacagaccgaacatcgtctacggaagccatagcaagatgcagagtgtt 734
 QY 288 cactaatgagcgtctcattgcttaacatttgggactcttctgattttctgacttt 347
 DB 735 ttcaaatgagcgtctcattgcttaacatttgggactcttctgattttctgacttt 794
 QY 348 aggagcttctaattggttaactatagagaaactctacagcgttcaatctggttattt 407
 DB 795 aggggactcaaatggatactcaaaagcaagttcggctgattcaacttgggttatt 854
 QY 408 cggagttaaggt-----actactgtaaatgcaaatgaactacccaacgtttc 455
 DB 855 aggggtttcagctgcaagctcaactctcaccgactcttccaatgcaacttccctaagcag 914
 QY 456 ttaagtaacgaggtgttgaactttacacagacactcttctcttggagcgtagcgc 515
 DB 915 cattaccacaggtgttggatattttatcacagacacatcatcttcttggagcgtagcgc 974
 QY 516 tctgagagccttatgggaatcggtgtgcaacttttggagcgtgaattccaatgacaca 575
 DB 975 acgtgagccttatgggaatgggtgtgcaacttttggagcgtgaattccaatgacaca 1034
 QY 576 gtccaaacctaagtggaagacttaattgtatctgtaacgtatcgcaattctctgtaa 635
 DB 1035 atctaactcagattgagatgctcaacgtcacttcaagccagacaaatttctgattca 1094
 QY 636 caaacccaagggtctaaaggcgtt-----gcttcccttgcgaacagacgctggcgt 689
 DB 1095 caaaccaagggtctaaaggcgttagctcgaaatttcttccactataacggtggaac 1154
 QY 690 acaacagctactggaacaaatctgcgaccatcaattatcaatgaatgcaagtaggagc 749
 DB 1155 acaagagctacagacacaaatcagctacaatttaataccatgaatggcagtaggacct 1214
 QY 750 ctctctattctacagactaaactctttagtgcatacattgagtagtacaattgctcagc 809
 DB 1215 cgccctgtctacagattgaatgctgttccatataattggcgttaactggtcaagac 1274
 QY 810 aacttttgatgctgataacatccgacttgcagcgaacaaactacacagctgttttaa 869
 DB 1275 aacttttgatgctgataacatccgacttgcagcgaacaaactacacagctgttttaa 1334
 QY 870 cttaactgcatggaaccccttcttactaggaatgccaagcattgtctactact----- 924
 DB 1335 cattactacatggaacccaagccttataggatacaacactgcttttgcgaataatgagtg 1394
 QY 925 ----gattcgtctcagactcgaatgcaaatgtttctgcatcagatacaacagtttaactc 980
 DB 1395 taagaggtctctatcgtatgtcttgcaaatgtcttcgattcagatacaacaaatgaagtc 1454
 QY 981 tagaaaagttgtgaggttactgtagagactcatttagttagtgcgtgataaatggtcaat 1040
 DB 1455 tagaaaagttgtgaggttactgtgagactcatttagttagtgcgtgataaatggtcaat 1514
 QY 1041 tactgcagagctcgttttaataacgagagagcgtctcagctatcgtggtcagttcagatt 1100
 DB 1515 cactggtgaagcagcgttaatacaatgaagagagcgtctcagctatcagattgcaattccgctt 1574
 QY 1101 c 1101
 DB 1575 c 1575

RESULT 5
 AAH56267
 ID AAH56267 standard; DNA; 1578 BP.
 AC AAH56267;
 DT 05-SEP-2001 (first entry)
 DE Chlamydia trachomatis pmp gene sequence.
 KW Chlamydia; vaccine; infection; fusion protein; antigen;
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
 KW acute respiratory tract infection; Cap1; CT529; OMCB;
 KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.
 OS Chlamydia trachomatis.
 XX WO200140474-A2.
 XX 07-JUN-2001.
 XX 04-DEC-2000; 2000WO-US32919.
 XX 03-DEC-1999; 99US-0454684.
 XX 19-APR-2000; 2000US-0556877.
 XX 20-JUN-2000; 2000US-0598419.
 PA (CORI-) CORIXA CORP.
 XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;
 WIPI; 2001-374831/39.
 XX Chlamydia polypeptides and fusion proteins useful for preventing pelvic
 inflammatory disease, trachoma, acute respiratory tract infections,
 atherosclerosis and heart disease -
 Claim 1; Page 211; 295pp; English.
 CC The present nucleotide sequence is provided in a specification
 relating to compounds and methods for the treatment and diagnosis of
 chlamydial infection. The compounds provided include polypeptides and
 fusion proteins comprising immunogenic portions of Chlamydia antigens
 and DNA sequences encoding such polypeptides. They are useful for
 vaccinating against chlamydial infection, which causes pelvic
 inflammatory disease, trachoma, acute respiratory tract infections,
 atherosclerosis and heart disease.
 XX Sequence 1578 BP: 414 A; 394 C; 382 G; 388 T; 0 other;
 Query Match 46.7%; Score 514.6; DB 22; Length 1578;
 Best Local Similarity 68.8%; Pred No. 9e-147;
 Matches 785; Conservative 0; Mismatches 314; Indels 42; Gaps 4;
 QY 3 gttgctgtagggaaccccttctgacacgttatttaattggtgacataatggaagg 62
 DB 435 gctgctgtggggaatcccaagtgaaacaaagtttatttaacgtgcactatggaagg 494
 QY 63 tctgcagagagatccttgcgactccttgcgactctggtgcgacgtattagcttgcgc 122
 DB 495 tcttcagagagatccttgcgactccttgcgactctggtgcgacgtattagcttgcgc 554
 QY 123 tggattttacgagactatgttttcgacgtatctttaaagtagatgcacctaataacatt 182
 DB 555 aggtactacgagagattatgttttcgactgttattaaagattgtagtaataaaacttt 614
 QY 183 ttctatggagcgaacgactactggtatccgctgctgcaaaactataactact----- 231
 DB 615 tagcgcatggtgcgaactcctacgcaggtctatagtaacgcaagtaataactacagcc 674
 QY 232 ----gcgtatagatagacttaacccgctcctacaataagaatttacagatgcagagtgtt 287

10

DE chlamydia trachomatis major outer membrane protein pNA:

Accession	Sequence	Length
Qy	5 tgctctgagggaaccctctctgatccaaagcttattaatctgattggtacaaatatgggaaggtg	64
Db	1355 tgctctgagggaatctctgtgcaacaaagcttatgatcgcgggaattcttatgggaaggtt	1414
Qy	65 ctgcaggagataccttgcgactcttgctgacttggatggcgcgtattagcttacgctctg	124
Db	1415 tcagcagaadactcttcgactcttcaccacacttgatgtagcgtcatcagcatcggtatgg	1474

QY 125 gatattacgagactatgttttcgaccgtatctttaaagtagatgcacctaataacatttt 184
 Db 1475 gttactatgtgtaactgtttgtttcgaccgtgttttgcaacagatgtgataaagaattcc 1534
 QY 185 ctatggagccaaagccta-----ctggatccgctgctgcaaacatactactacgcgtag 238
 Db 1535 aaatgggtgcaagctcaacactgtacaggaatgtgctgagctccatcttgcattgacag 1594
 QY 239 atagactaacccgctcaataagcattacacgagtcacagtggtttcactaatagag 298
 Db 1595 caagagagaatcctgttccagccagacatagcagagtgctgagatgtttacaaatgctg 1654
 QY 299 gcttcattccttaaacatttgggagcgttttgatgttttctgtacttttaggagcttcta 358
 Db 1655 ctacatggtcattgaatttgggagcgttttgatgtattctgtacattaggagccacca 1714
 QY 359 atggttacattagagaaactctacagcgttcaactctggttttatttcggagttta--- 415
 Db 1715 gggatattcttaaggaattcagcatcttcaacttagttggtcttattcggagataatg 1774
 QY 416 -----aaggtagctctgtaaatgcaaatgaactacccaacgttttcttaagtacggag 469
 Db 1775 agaaccatgctacagtttcagatagtaagcttgtaaccaaatagcttagatcaatctg 1834
 QY 470 ttgtgaactttacacagacaccttttctgtgagcgttaggcgtgtgagccttat 529
 Db 1835 ttgtgagttgtatacagatactatttctgtgagcgttaggcgtgtgagccttat 1894
 QY 530 gggaaatcggttgcacacttggagcgtgaattcaataatcacagtcacaaacctaaag 589
 Db 1895 gggaaatcggttgcacacttggagcgtgaattcaataatcacagtcacaaacctaaag 1954
 QY 590 ttgaagaacttaattgtagctgtaacgtatcgtaactctctgttaacaaacccaaaggct 649
 Db 1955 tcgaagaattaaacgtttctgttaacgagcgtgagtttactatcaataagcctaaaggat 2014
 QY 650 ataaagcgttcttcccttgcacacagcgtggtgagtagcaacagctactggaacaa 709
 Db 2015 atgtagggcaagaattccctctgtatcttaaaagcaggaacagatggtgtgacaggaacta 2074
 QY 710 agtctgcgaccatcaattatcatgaatgcaagtaggagcctctctctctacagactaa 769
 Db 2075 agatgctctctattgattaccatgaatggcaggaagtttagctctctctacagactga 2134
 QY 770 actctttagtgcacatactgagtagcaatggtctcgagcaactttttagctgataaca 829
 Db 2135 atatgttcaactccctacattggagttaaatggtctcgagcaagttttagctgacagacga 2194
 QY 830 tcgcaattgctcagcaaaactacacagctgttttaacttaactgcatgcaagaccctt 889
 Db 2195 tctgattgtcagcgaagtcagctacacagctgtttttagctgtattaccactctgaacccaa 2254
 QY 890 ctttactaggaatgccacagcattgtctactactga---tctgtctcagacttcaatgc 946
 Db 2255 ctatgctgagcgtgagcgtgaaagctagcgcagaggtgagctgcgagatacatatgc 2314
 QY 947 aaattgttctgtcagatcaacaagtttaaatctagaaaaagctgtgagttactgtag 1006
 Db 2315 aaatggttctgtgaattgaaacagatgaaatctagaaaaatcttgcggtattgtagtag 2374
 QY 1007 gagctactttagtgaatggtgagcagcagcagcagcagcagcagcagcagcagcagc 1066
 Db 2375 gaaacactattgtgagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 2434
 QY 1067 agagagctgctcagctatctgtgttcagtttcagattc 1101
 Db 2435 agagagctgctcagctatctgtgttcagtttcagattc 2469

RESULT 7
 AA292753
 ID AA292753 standard; DNA; 3133 BP.

XX AA292753;
 AC 19-JUN-2000 (first entry)
 DE DNA encoding Chlamydia trachomatis major outer membrane protein (MOMP).
 KW Major outer membrane protein; MOMP; serovar L2; immunoreactive;
 KW antibody production; immunoassay; detection; vaccine; trachoma;
 KW inclusion conjunctivitis; pneumonia; lymphogranuloma venereum;
 KW mucous membrane genital tract infections; ds.
 OS Chlamydia trachomatis serovar L2.
 XX Key Location/Qualifiers
 PH 1288..2472
 FT /*tag--a
 FT /product= "Chlamydia trachomatis serovar L2 MOMP"
 XX US6030799-A.
 XX 29-FEB-2000.
 XX 06-JUN-1995; 95US-0466152.
 XX 28-OCT-1993; 93US-0144095.
 PR 13-JAN-1986; 86US-0818523.
 PR 25-APR-1991; 91US-0691639.
 PR 14-JAN-1985; 85US-0692001.
 XX (WASH-) WASHINGTON RES FOUND.
 PI Agabian N, Stephens R, Kuo C-C, Mullenbach G;
 DR WPI: 2000-223163/19.
 DR P-PSDB; AA81268.
 XX New synthetic or recombinant polypeptide, useful for diagnosing or
 PT preventing Chlamydia trachomatis infection, is immunologically
 PT equivalent to a major outer membrane protein
 XX Examples; Fig 2; 17pp; English.
 CC The present sequence represents DNA encoding the major outer membrane
 CC protein (MOMP) of Chlamydia trachomatis serovar L2. DNA obtained from
 CC C. trachomatis serovar L2 was partially digested with DNase I and
 CC inserted into the bacteriophage vector lambda-gt11; a lambda-gt11 insert
 CC (AA292752; ATCC #40157, referred to as lambda-gt11/L2/33 in the
 CC specification) was obtained which encodes a protein fragment (AA81267)
 CC reactive with a pool of C. trachomatis-specific monoclonal antibodies.
 CC The lambda-gt11/L2/33 insert sequence was also used to probe a
 CC Chlamydia genomic DNA phage lambda 1059 library, and the present
 CC sequence which encodes the full-length Chlamydia trachomatis serovar L2
 CC MOMP was identified. C. trachomatis is a human pathogen responsible for
 CC diseases such as trachoma, inclusion conjunctivitis, pneumonia,
 CC lymphogranuloma venereum, and mucous membrane genital tract infections
 CC such as cervicitis and urethritis. The full-length recombinant MOMP or a
 CC fragment thereof is used to elicit the production of antibodies to a MOMP
 CC of Chlamydia trachomatis. MOMP proteins are also useful as immunoassay
 CC reagents (substitutes for native MOMP) for detecting C. trachomatis or
 CC its antibodies, for diagnosing infection, or as an immunogen for
 CC vaccines.
 XX Sequence 3133 BP; 911 A; 670 C; 609 G; 943 T; 0 other;

Query Match 43.5%; Score 479.4; DB 21; Length 3133;
 Best Local Similarity 66.5%; Pred. No. 6.9e-136;
 Matches 741; Conservative 0; Mismatches 356; Indels 18; Gaps 3;

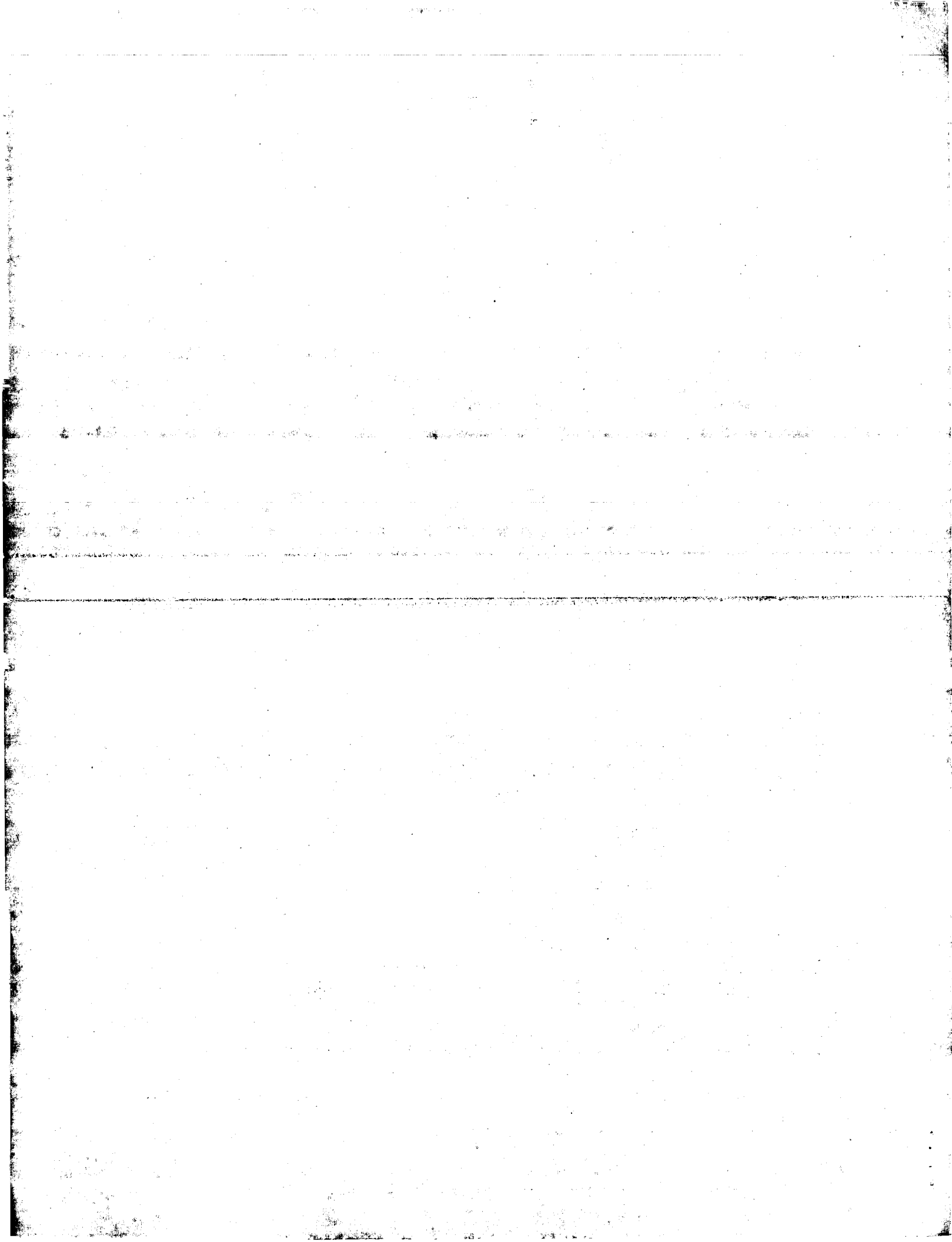
QY 5 tgcctgtagggaacccttctgacacagcttattattgattggtacaaatggaaggtg 64
 Db 1355 tgcctgtagggaacccttctgacacagcttattattgattggtacaaatggaaggtg 1414

Qy	65	ctgcaggagatcctctgcgatacctctgcgtactctgcgtactcttggtgcgacgctattagcttaccgtgctg	124
Db	1415	tcggcgagatcctctgcgatacctctgcacacattggtgtgcgctatcagatcgcgtatgg	1474
Qy	125	gattttacggagagactatgttttcgacgctatccttaaaagtagatgcacctaaacatttt	184
Db	1475	gttacctatggtgactttgttttcgacgctgttttgcaaacagatgtgaaataagaattcc	1534
Qy	185	ctatggagcgcaagccta-----ctggatccgctgctgcgaacctactactactgcgctg	238
Db	1535	aaatgggtgccaaagcctacaactgctacagcgaatgctgcagctccatccactgtgtacag	1594
Qy	239	atagacctaaocggcctacaataagcatctttacacgatgcagatgcagatggttccactaatgcag	298
Db	1595	caagagagaatcctgctttacggcgacatatgcagatgctgagatgttttacaatgctg	1654
Qy	299	gcttcattgccttaaacatttgggacgcttttgatgtttttctactttttaggagcttcta	358
Db	1655	cttcatggcattgaatatctgggacgcttttgatgtattctgacatttgggagcca	1714
Qy	359	atggttaccattagaggaactctacagcgttccatctcgtttgttttattcoggagttta---	415
Db	1715	gtggatatottaaaggaaattcagcatctttcaacttagttggcttattcoggagataatg	1774
Qy	416	-----aagttactactgtaaatgcaaatgaactaccaaagtttctttaagtaacggag	469
Db	1775	agaacctgctacagtttcagatagtaagctgtgtaccaaatatgagcttaccatctg	1834
Qy	470	tgttgaaactttacacagacacctttctcttgagcgttaggcgctctgtagccttat	529
Db	1835	tgttgagttgtatacagatactatttctggtgagctgtggagctgtgcagctttgt	1894
Qy	530	gggaatgcggttgtgcaactttgggagctgaattccaatatgcacagtccaacctaaag	589
Db	1895	gggaatgggatgcgacgttttagcgtcttcttccaatacgcctcaatccaagcctaaag	1954
Qy	590	ttagaagacttaatgtgactgttaacgtatgcgaattctctgttaacaaacccaagggct	649
Db	1955	tcgaagaattaaacgttctgttaacgcgctgtgatttactatcaataagcctaaagat	2014
Qy	650	ataaaggcgtgtgtttcccttgcgaacagacgctggcgttagcaacagctactggaaaca	709
Db	2015	atgtaggcgaaagtccctcttgcatttaaagcaggaaacagatggtgtgacaggaaact	2074
Qy	710	agctcgcacacataattatcatgaatggcaagttaggcctctctacttccagactaa	769
Db	2075	aggaagcctctatgtattaccatgaatggcgaagcttttagctctctcttccagactga	2134
Qy	770	actctttagtgccatacttgagtcacaaatggtctcgagcaacttttgcgtgtataaca	829
Db	2135	atatgttcactccctacattggagttaaatggtctcgagcaagttttgatgcagacga	2194
Qy	830	tcgcgattgtcgaagcaaaactactacagctgtttttaaacttaactgcagcgaacctt	889
Db	2195	ttcgtattgtcgcgcgaagtcagctacacgtcttctgtgttaccactctgaaccca	2254
Qy	890	ctttactaggaatgcacagcattgtctactactga-----ttcgttctcagacttcatgc	946
Db	2255	ctattgctggagctggcgatgtgaaagcttagcgcagaggggtcagctcggagataccatgc	2314
Qy	947	aaattgtttcctgtcagatacaacaagtttaaatctagaaaaagccttgcggagtactgtag	1006
Db	2315	aaatcgtttccttgcaattgaacaagatgaatctcagaaaatcttgcggtatttcagtag	2374
Qy	1007	gagctactttagttgatgctgataaatggtcacttacttgcagagaagctcgtttaataacg	1066
Db	2375	gaacaaactatgtggatgcagacaaataacgcaggttacagttgagactcgtgtgatcagtg	2434
Qy	1067	agagagctctcagctatcttgcgttccagtttcagattc	1101
Db	2435	aqacagctgtctcagctaaatgcacaaattccgcttc	2469

RESULT	8	
ID	AAV40646	
XX	AAV40646 standard; DNA; 3133 BP.	
XX		
AC	AAV40646;	
XX		
DT	27-OCT-1998 (first entry)	
XX		
DE	Chlamydia trachomatis serovar L2 MOMP coding region.	
XX		
KW	MOMP; major outer membrane protein; immunoassay; diagnosis;	
KW	detection; antibody; serovar L2; ds.	
XX		
OS	Chlamydia trachomatis.	
XX		
PH	Key	Location/Qualifiers
FT	CDS	1288..2472
FT		/*tag= a
FT		/product= MOMP
FT	sig_peptide	1287..1353
FT		/*tag= b
XX		
PN	US5770714-A.	
XX		
PD	23-JUN-1998.	
XX		
PF	14-JAN-1985; 85US-0692001.	
XX		
PR	13-JAN-1986; 86US-0818523.	
PR	14-JAN-1985; 85US-0692001.	
PR	25-APR-1991; 91US-0691639.	
PR	28-OCT-1993; 93US-0144095.	
PR	06-JUN-1995; 95US-0466814.	
XX		
PA	(CHIR) CHIRON CORP.	
PA	(WASH-) WASHINGTON RES FOUND.	
XX		
PI	Agabian N, Kuo C, Mullenbach G, Stephens R;	
XX		
DR	WPI; 1998-376887/32.	
DR	P-PSDB; AAW57775.	
XX		
PT	DNA coding for Chlamydia trachomatis polypeptide - useful for	
PT	producing recombinant polypeptide, etc.	
XX		
PS	Claim 5; Fig 2; 15pp; English.	
XX		
CC	The sequence is that encoding a major outer membrane	
CC	protein (MOMP) of Chlamydia trachomatis. This polypeptide can	
CC	be used in immunoassays, e.g. to detect Chlamydia trachomatis	
CC	antibodies in blood, or can be used in vaccines. The	
CC	polynucleotide can be labelled and used as a diagnostic probe.	
XX		
SQ	Sequence 3133 BP; 909 A; 670 C; 610 G; 944 T; 0 other;	

	Query Match	43.4%	Score 477.8;	DB 19;	Length 3133;
	Best Local Similarity	66.4%;	Pred. No. 2.le-135;		
	Matches 740;	Conservative 0;	Mismatches 357;	Indels 13;	Gaps 3;
<hr/>					
Qy	5	tgcctgtagggaaccccttcgatccaagcttattaaattgattggtacaatatgggaagtgg	64		
Dd	1355	tgctctgtgggaatcctcgctaaccgaaccttatgatcgacggaattctatgggaaggtt	1414		
Qy	65	ctgcaggagataccttgcgataccttgcgcactttgctgcgcgctatttagctttacgttgctg	124		
Dd	1415	tcggcggagataccttgcgataccttgcaccaacttgtgtgcgcgtatcagcatgcgtatgg	1474		
Qy	125	gattttaacggagactatattttogaaccttatctttaaagtagtagatgcacctaaaacatttt	184		
Dd	1475	gttaactatggtagctattgttttcgacctgttttgcacaacagatgtgaataagaagattcc	1534		

Search completed: February 7, 2002, 20:23:45
Job time: 16539 sec



XX A soluble fused protein useful for diagnosis of Chlamydia infection,
 PT comprises at least part of major outer membrane protein (MOMP) of
 PT Chlamydia trachomatis ;
 XX
 PS Claim 16; Page 28; 37pp; Japanese.
 XX
 CC The present invention describes fusion proteins (I) comprising at least
 CC part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
 CC at least one hydrophilic polypeptide having no immunoreactivity to
 CC human serum and their connected part. AAA08120 to AAA08125 encode
 CC specifically claimed examples of the fusion proteins given in AY82388
 CC to AY82393. Also described is a method (A) for the detection of
 CC Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for
 CC the diagnosis of Chlamydia trachomatis infectious diseases. The method
 CC can diagnose Chlamydia trachomatis infectious diseases specifically in a
 CC high sensitivity.
 XX
 SQ Sequence 1362 BP; 383 A; 283 C; 324 G; 372 T; 0 other;

Query Match 26.1%; Score 287.4; DB 21; Length 1362;
 Best Local Similarity 66.6%; Pred. No. 1.8e-77;
 Matches 427; Conservative 0; Mismatches 211; Indels 3; Gaps 1;

QY 464 acggaggtgtgaactttacacagacacctcttctctggagcgtgagcgtctgtgag 523
 DB 722 acgaaattgtagaaggtatcacagatactacttttctgtggagcgtgagcgtgag 781
 QY 524 ccttattggaatgcggtgtgcaactttggagcgtgaaattccaatgcacagtcacaaac 583
 DB 782 ctttgcgggaatgtaggcgagcgtttaggcgtctcttcccaatcacgctcaatcaagc 841
 QY 584 cttaaagtgaagaacttaattgatctgttaacgtatcgcaattctctgttaacaaaccca 643
 DB 842 cttaaagtcgaagaattaaactttctgttaacgcagcgtgagttactatcaataagccta 901
 QY 644 agggctataaaggcgtgtcttcccttgcacacgaagcgtgagcgtgagcgtgagcgtg 703
 DB 902 aaggatattgtaggaagaattccctctgtgattcttaagcaggaacagatggtgtgacag 961
 QY 704 gaacaaagctgcgaccatcaattatcatgaatggaatggaagtaggagcgtctctatcttaca 763
 DB 962 gaactaaggatgcctctattgattaccatgaatggaagcaggaagtttagctctctcttaca 1021
 QY 764 gactaaactctttagtgcatacatgagtagtacaatggtctgcgcaactttttagctgtg 823
 DB 1022 gactgaatatgttcaactcctcattgagtttaaatggtctgcgcaagttttagctgag 1081
 QY 824 ataacatccgcatgtctcagcacaactacacacgtgttttaaacattacgtcatgga 883
 DB 1082 acacgattcgtattgtcagcagagcgtgagcgtacacacgtcttctgtgattaccactctga 1141
 QY 884 accctctcttactaggaatgccagcagcattgtctactactga---ttcgtctctcagact 940
 DB 1142 acccaactattgtagcgtgagcgtgagtgaaagcgtgagcagaggtgagcgtcggagata 1201
 QY 941 tcatgcaaatgttccctgtcagatcaacaaagtttaaatcttagaataaagcgttgcgagta 1000
 DB 1202 ccatgcaaatgttccctgtcagatgaaagcgtgagcgtgagcgtgagcgtgagcgtgag 1261
 QY 1001 ctgtaggagcactttagtgcgtgataaaatggtcactactcagcagagcgtgagcgtgag 1060
 DB 1262 cagttaggaacaaactattgtagcagcagcaataacgcagcgtgagcgtgagcgtgag 1321
 QY 1061 ttaacgagagagcgtcagcgtatctgtgtcagttcagattc 1101
 DB 1322 tegatgagagcgtcagcgttaaatgcaaatctcgtcttc 1362

RESULT 15
 AAA08120
 ID AAA08120 standard; DNA; 1371 BP.

XX AAA08120;
 AC
 XX
 DT 27-JUN-2000 (first entry)
 XX
 DE
 XX C. trachomatis MOMP containing fusion protein nucleotide sequence #1.
 KW Chlamydia trachomatis; fusion protein; major outer membrane protein;
 KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;
 KW infection; infectious disease; ds.
 XX
 OS Chlamydia trachomatis.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1371
 FT /tag= a
 FT /note= "Fusion protein containing at least part of a
 FT major outer membrane protein (MOMP) of Chlamydia
 FT trachomatis; no stop codon given"

JP2000041678-A.
 15-FEB-2000.
 28-JUL-1998; 98JP-0213212.
 28-JUL-1998; 98JP-0213212.
 (ELED) DENKI KAGAKU KOGYO KK.
 WPI; 2000-295780/26.
 P-PSDB; AAY82388.
 A soluble fused protein useful for diagnosis of Chlamydia infection,
 comprises at least part of major outer membrane protein (MOMP) of
 Chlamydia trachomatis -
 Claim 12; Page 26-27; 37pp; Japanese.
 The present invention describes fusion proteins (I) comprising at least
 part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
 at least one hydrophilic polypeptide having no immunoreactivity to
 human serum and their connected part. AAA08120 to AAA08125 encode
 specifically claimed examples of the fusion proteins given in AAY82388
 to AAY82393. Also described is a method (A) for the detection of
 Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for
 the diagnosis of Chlamydia trachomatis infectious diseases. The method
 can diagnose Chlamydia trachomatis infectious diseases specifically in a
 high sensitivity.
 Sequence 1371 BP; 390 A; 276 C; 324 G; 381 T; 0 other;

Query Match 26.1%; Score 287.4; DB 21; Length 1371;
 Best Local Similarity 66.6%; Pred. No. 1.8e-77;
 Matches 427; Conservative 0; Mismatches 211; Indels 3; Gaps 1;
 QY 464 acggaggtgtgaactttacacagacacctcttcttggagcgtgagcgtctgtgag 523
 DB 731 acgaaattgtagaaggtatcacagatactacttttctgtggagcgtgagcgtgagcgtgag 790
 QY 524 ccttattggaatgcggtgtgcaactttggagcgtgaaattccaatgcacagtcacaaac 583
 DB 791 ctttgcgggaatgtaggcgagcgtttaggcgtcttcttccaatcacgctcaatccaaacg 850
 QY 584 cttaaagtgaagaacttaattgatctgataacgtatcgcaatctctctgataacaaaccca 643
 DB 851 cttaaagtcgaagaattaaacgtctctgtcgaacgcagcgtgagttactatcaataagccta 910
 QY 644 agggctataaaggcgtgtcttcttcccttgcacacagcgtgagcgtgagcgtgagcgtgag 703
 DB 911 aagatatgtagggcgaagaattccctctgtatcttaagcaggaacagatggtgtgacag 970

[illegible]

Db 616 tggatcaatcactggtgaagcagcgttaataatcaatgaagagcgctcaacatgaatgctcaa 675
 QY 1093 ttcagattc 1101
 Db 676 ttcagattc 684

RESULT 12

AAA08124
 ID AAA08124 standard; DNA; 1047 BP.

XX
 AC AAA08124;

XX
 DT 27-JUN-2000 (first entry)
 XX

DE C. trachomatis MOMP containing fusion protein nucleotide sequence #5.
 KW Chlamydia trachomatis; fusion protein; major outer membrane protein;
 KW MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;
 KW infection; infectious disease; ds.

XX Chlamydia trachomatis.
 OS Synthetic.

XX Location/Qualifiers

XX Key 1..1047

XX CDS /*tag= a

XX FT /note= "Fusion protein containing at least part of a
 major outer membrane protein (MOMP) of Chlamydia
 trachomatis; no stop codon given"

XX JP2000041678-A.

PN 15-FEB-2000.

PD 28-JUL-1998; 98JP-0213212.

PR 28-JUL-1998; 98JP-0213212.

XX (ELED) DENKI KAGAKU KOGYO KK.

XX WPI; 2000-295780/26.

DR P-PSDB; AAY82392.

XX A soluble fused protein useful for diagnosis of Chlamydia infection,
 PT comprises at least part of major outer membrane protein (MOMP) of
 PT Chlamydia trachomatis

XX Claim 20; Page 29-30; 37pp; Japanese.

CC The present invention describes fusion proteins (I) comprising at least
 CC part of a major outer membrane protein (MOMP) of Chlamydia trachomatis,
 CC at least one hydrophilic polypeptide having no immunoreactivity to
 CC human serum and their connected part. AAA08120 to AAA08125 encode
 CC specifically claimed examples of the fusion proteins given in AAY82388
 CC to AAY82393. Also described is a method (A) for the detection of
 CC Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for
 CC the diagnosis of Chlamydia trachomatis infectious diseases. The method
 CC can diagnose Chlamydia trachomatis infectious diseases specifically in a
 CC high sensitivity.

XX Sequence 1047 BP; 315 A; 209 C; 233 G; 290 T; 0 other;

Query Match 29.9%; Score 328.8; DB 21; Length 1047;
 Best Local Similarity 60.5%; Pred. NO. 3.7e-90;
 Matches 671; Conservative 0; Mismatches 367; Indels 72; Gaps 5;

QY 1 atgttgcctgttaggaacaccttctgatacgaacttaattgattggtacacatgggaa 60

Db 1 atgttgcctgttggtaaacctgtgtaacacaccccttatgatacgcaggtacctatgggaa 60

QY 61 gdtgtgcaggagatccttgcgatacttgcgctacttggcgagcgtattagcttactg 120

Db 61 gdtgtgcaggagatccttgcgatacttgcgctacttggcgagcgtattagcttactg 120
 QY 121 gctgtgatttttgcagagactatgttttcacgcgtatcttataaagttagaccctaaaca 180
 Db 121 atgggttactatggtgacttgttttcacgcgtatcttataaagttagaccctaaaca 180
 QY 181 ttttctatggagccaagccta-----ctggatccgctgctgcacaaacttactactgcc 234
 Db 181 ttcacaaatgggtgccagcctacaactgtacaggcaatgctgcagctccatccacttgt 240
 QY 235 gtatagatagacctaacccgcgtacataaagaattttacacgatagcagagtggttcaactaa 294
 Db 241 acagcaagagagaatcctgcttaacgcgcgacatgcatgaggatgctgagatgtttacaat 300
 QY 295 gtaggtctcaattgctttaaacaatttggatgcttggatgcttggatgcttggatgcttggatgct 354
 Db 301 gctgttcatagcgaattgaataatttggatgcttggatgcttggatgcttggatgcttggatgct 360
 QY 355 tctaattggttacaattagaggaactctacagcgttcaatctcgttgggtttatttcggagtt 414
 Db 361 accagtgatattctaaaggaaattcagcattcttcaacttagttggttatttcggagat 420
 QY 415 aaagttactactgtaaatgcaaatgaatacacaacgctttctttaaagtaacggagttgtt 474
 Db 421 aatg-----agaaccatgctacagtttcagatagtaagcttgcacca 462
 QY 475 gaactttcacacagacaccttcttcttggagcgtaggcgtggtggagccttaaggaa 534
 Db 463 aatagagcttagatcaatct----- 483

QY 535 tgcggtgtgcacacttgggagctgaattccaatatagcacagtcacaaacctaaagttaa 594
 Db 484 ----gtgtgagttgtatgtaacaataacaataacaataacaataacaataacaata 539
 QY 595 gaacttaattgactgttaacgctatgcgaattctctgttaacaaacccaagggctataaa 654
 Db 540 caa--taacaataacaataacaataagctgagtttactatacaataagcctcaaggatgata 597
 QY 655 ggcgtgttcttcccttgcacacagcgtggtgtagcaacagctactctggaacaaagtct 714
 Db 598 gggcaagaattccctcttgatcttaagcaggaacagatggtgtgacaggaactaaggat 657
 QY 715 gcgacactcaattatcatgaatgagcaagtaggagcgtctctctattctacagactaaactct 774
 Db 658 gcctctattgattacatgaatgagcaagcaagtttagctctctctctacagactgaatg 717
 QY 775 ttagtgcacattgagtagtaacaatggtctcgaacaaacttctgagtagtaacacacgc 834
 Db 718 ttaactccctcattggagtttaaatggctcgcagcaagtttctgtagcagacacgattcgt 777
 QY 835 attgctcagccaaactacacagctgtttttaaacttaactgcatggaacacctcttcta 894
 Db 778 attgctcagcgaagtcagctacacactgtcttctgattacacactctgaacccaactatt 837
 QY 895 ctaggaaatgccacagaattgtctactactga---ttcgttctcagacttcatgcaaat 951
 Db 838 gctggagctgcgctggtgaaagctagcgcagaggtgcagctcggagctaccatgcaatc 897
 QY 952 gtttctctcagatcaacaagtttaaatctagaaaaagctgtgtggagttactgttagagct 1011
 Db 898 gtttctctgcaattgacaagaatgaataatctagaaaaattctgcggtattgcagtaggaaca 957
 QY 1012 acttttagttgctgataaaatggtcacttactgcagaagctgttatttaataacagagaga 1071
 Db 958 actattgtggtgagacacaaatacagcagttacagttgagactcgtgtgatcgatgagaga 1017
 QY 1072 gctgctcagctatctgttgcagttcagattc 1101
 Db 1018 gctgctcagtaaatgcacaattccgcttc 1047

RESULT 13

299 gcttcattgccttaaacattttqqqatcgcctttgatgttttctgtactttaggagcctcta 358

Query Match	99.7%	Score 1098;	DB 1;	Length 1830;
Best Local Similarity	100.0%;	Pred. No. 1.le-275;		
Matches 1098; Conservative	0;	Mismatches	0;	Gaps
QY	4	ttgcctgagggaaccccttcgatccaaagcttatttaattgatgtgtacaatatgggaaggt	63	
Db	386	TTGCCGTAGGAACCCCTCTGATCCAAAGCTTATTATTGATGGTACAAATATGGAAAGGT	445	
QY	64	gtcgcaggagatccttgcgataccttgcgctactcttgctgcagcgtctattagcttacygt	123	
Db	446	GCTGCAAGAGATCCTTGCGATCCTGGCTACTGTGTGCGACGCTATTAGCTTACGTGCT	505	
QY	124	ggattttcaggagactatgttttcgacccgttatcttaaaagtagatgcacctaacaacattt	183	
Db	506	GGATTTTACGGAGACTATGTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT	565	
QY	184	tctatgggagccaagcctactgataccgctgctgcgaaactatactactgcgtagataga	243	
Db	566	TCATGGGAGCCAAGCCTACTGTGATCGCTGCTGCACCACTATATACTAGCGGTAGATAGA	625	
QY	244	cctaaccggcctcacataaagcatttacacgatgcagatggttccactaatgcaggttc	303	
Db	626	CCTAACCCGGCCTACAAATAGCATTTACACGATGCAGATGGTTCACTAATGCAGGCTTC	685	
QY	304	atgctcttaaacatttgggactgcctttgatgtttctctgtactcttaggagctcttaaggt	363	
Db	686	ATTGCCCTTAAACATTTGGGATCGCTTGTATGTTTCTGTACTTATGAGAGCTTCTTAATGGT	745	
QY	364	tacattagaggaaactctacagcgttccaactcctggtgtttattcggagttaaaggtact	423	
Db	746	TACATTAGAGGAAACTCTACAGCGTCCAATCTCGTTGGTTTATTCGGAGTTAAAGGTACT	805	
QY	424	actgtaaatgcgaatgaactcaaacacggttctttaaagtaacggagttgttgaactttac	483	
Db	806	ACTGTAAATGCAATGAACCTACCAACGTTTCTTTAAGTAACGGAGTGTGTGAACCTTATC	865	
QY	484	acagacacctcttctcttggagcgttaggcgctcgtaggccttatgggaatgcggttgt	543	
Db	866	ACAGACACCTCTTCTCTTGGAGGTAGCGCTCGTGGAGCCCTTATGGGAATCGGTTGT	925	
QY	544	gcaactttgggagctgaattccaatatgcacagtcgcaaacctcaaaagttgaagaacttaac	603	
Db	926	GCAACTTTGGGAGCTGAATTCCTAAATATGCACAGTCCAAACCTTAAAGTTGAAGAACTTAAT	985	
QY	604	gtgatctgtaacgftatcgcaattctctgtaaacaaaccccaagggctataaagcgttgtct	663	
Db	986	GTGATCTGTAAACGTATCGCAATTTCTGTGTAAACAAACCCCAAGGGCTATAAAGCGTGTCT	1045	
QY	664	ttcccttgcacacagacgctggcgttagcaacagctactgtggaacaaagctcgcgacatc	723	
Db	1046	TTCCCTTGGCAACACAGCGTGGCGTAGCAACAGCTACTGGAACAAAGTCTGGGACCATC	1105	
QY	724	aattatcatgaatggcaagtagagcctctctatcttaccagactaaactcttagtgcca	783	
Db	1106	AATTATCATGAATGGCAAGTAGAGCGCTCTCTATCTTACAGACTAACTCTTTTAGTGCCA	1165	
QY	784	tacattggagtacaatgggtctcgacaaacttttgatgtgatacaacatccgactgtctcag	843	

12]	Shirai M., Hirakawa H., Ouchi K., Tabuchi M., Kishi F., Kimoto M.,	
RA	Takeuchi A., Nishida J., Shibata K., Fujinaga R., Yoneda H., Matsushima H.,	
RA	Tanaka C., Furukawa S., Miura K., Nakazawa A., Ishii K., Shiba T.,	
RA	Hattori M., Kuhara S.;	
RT	"Comparison of outer membrane protein genes omp and pmp in the whole genome	
RT	sequences of Chlamydia pneumoniae isolates from Japan and US";	
RL	Unpublished.	
XX		
XX	SWISS-PROT; P27455; OMP1_CHLPN.	
XX		
Key	Location/Qualifiers	
PH		
FT	1. .1170	
FT	/db_xref="taxon:83558"	
FT	/note="synonym:Chlamydia pneumoniae"	
FT	/sequenced_mol="DNA"	
FT	/organism="Chlamydia pneumoniae"	
FT	/strain="J138"	
FT	1. .1170	
FT	/codon_start=1	
FT	/db_xref="SWISS-PROT:P27455"	
FT	/transl_table=11	
FT	/gene="ompA"	
FT	/product="major outer membrane protein"	
FT	/protein_id="BAA85940.1"	
FT	/translation="MKKILKSAALSAAGFAGSVGSLQALFVGNPSDPSLLIDGTIWEGAA	
FT	GPDCPCATWCDAISLRAGFYGDYVDFRILKVDAPKTFESMGAKPTGSAANYTTAVDRP	
FT	NPAYNKLHDAEWFNTNAGFIALNIWDRDVFCTLCASNGYIRGNSAFNLVLGFVGKT	
FT	TYNANLPHVSLSGNVELYTDTSPFSWSVGARGALWEGCGCATLGAEFQYAOQSKPKVEL	
FT	NVICNVQSQSNKPKGYGVAEPPLTDAGYATATGKTSATINHEWQYAGSLSYRLSL	
FT	VPYIGVQWSRATFDADNIRIQAOKPLPTAVLNLTANNPILLGNATLSITDTSFSDPMQIV	
FT	SCQINKFKRKACGVTGVTALADKWSLTAEARLINERAAHVSGQFR"	
XX		
XX	Sequence 1170 BP: 320 A; 255 C; 246 G; 349 T; 0 other;	
XX		

Query Match	99.7%	Score 1098	DB 15	Length 1170
Best Local Similarity	100.0%	Pred. No. 1.le-275		
Matches 1098	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	4	ttgcgtgtaggaacccctctgaccgaagcttattaatgtagtgatacaatatgggaaggt	63	
DB	70	TTGCCTGTAGGAACCCCTCTGATCCAAAGCTTATTAAATTGATGGTACAATATGCGAAGGT	129	
QY	64	gctcaggagagacccctgacatcccttgcgtaacttggtgcgaacctattagcttaccgtgt	123	
DB	130	GCTCGAGAGATCCTGCGATCCTTGGCGCTACTTGGTGGCAGCCTATTAGCTTACCTGGCT	189	
QY	124	ggattttacgagactatgttttcgacgcatctcttctaaagttagatgcacctaataacattt	183	
DB	190	GGATTTTACGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT	249	
QY	184	totatgggagccagcctaactggaatccgctgctgcgaacctatactactgcgtagataga	243	
DB	250	TCTATGGGAGCAAGCCTACTTGGATCGCGTCTGCAAACTATACTACTGCGGTAGATAGA	309	
QY	244	ccataccgcgcctacaataaagcatttacagatgcagatgggttcactaatgcaggcttc	303	
DB	310	CCTAACCCGGCCTACAATAAGCATTTACAGATGCAGATGGTTTCACTAATGCAGGCTTC	369	
QY	304	attgccttaaacatttgggatcgtgttgatgttttctgtacttttaggagcttcaatgggt	363	
DB	370	ATTGCGCTTTAAACATTTGGGATCGCTTTGATGTTTCTGTACTTTTGGAGGCTTCTTAATGGT	429	
QY	364	tacattagagaaacctctacagcgttcaactctggtttatttcggaggttaagggttact	423	
DB	430	TACATTAGAGGAACCTCTACAGCGTTCAATCTCGTGTGGTTTATTTCGGAGTTAAAGGTACT	489	
QY	424	actgtcaaatgcaaatgaaactccaaacgcttcttttaagtaacgaggtgtgtgaactttac	483	
DB	490	ACTGTAAATGCAAAATGAATCACTACCAACGTTTCTTTAAGTAACGAGGTGTGTGAACCTTAC	549	

```

RESULT      3
CTNMOPP
LOCUS      1830 bp      DNA      BCT
DEFINITION Chlamydia pneumoniae major outer membrane protein precursor,
            complete cds.
ACCESSION M69230
VERSION   M69230.1 GI:144540
KEYWORDS  major outer membrane protein.
SOURCE    Chlamydia pneumoniae DNA.
ORGANISM  Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydophila.
REFERENCE 1 (bases 1 to 1830)
AUTHORS   Perez Melgosa.M., Kuo,C.-C. and Campbell,L.A.
TITLE      Sequence analysis of the major outer membrane protein gene of
            Chlamydia pneumoniae
JOURNAL    Infect. Immun. 59, 2195-2199 (1991)
MEDLINE    91244474
FEATURES   Location/Qualifiers
            1..1830
            /organism="Chlamydophila pneumoniae"
            /db_xref="taxon:83558"
            306..310
            /note="putative"
            317..385
            317..1486
            /codon_start=1
            sig_peptide
            CDS
            /transl_table=1
            /protein_id="AAAF3071.1"

```

JOURNAL Submitted (26-FEB-1999) Pathology, Vanderbilt University, C-3321
 Medical Center North, Nashville, TN 37232, USA

FEATURES
 Source 1..1170
 /organism="Chlamydomonas pneumoniae"
 /db_xref="taxon:83558"
 /note="from the cerebral spinal fluid of a patient with rapidly progressive multiple sclerosis"
 1..1170
 /gene="MOMP"
 1..1170
 /gene="MOMP"
 /codon_start=1
 /transl_table=11
 /product="major outer membrane protein"
 /protein_id="AAD22492.1"
 /db_xref="GI:4545321"
 /translation="MKLLKSLLSAFAFSGVLSQALPVGNPSPDPSLLIDGTWEGA
 AGDPDPCATWCDALSLRAGFYGDYFEDRLKVDAPKTFSGAKPTGSAANYYTAVD
 RNPAYNKHLDHAEWFTNAGFIALNIMDFVCTLGASNGYIRGNSTAFNLVLFV
 KGTVMNHELNVSLNGVELYDTSFSSWVGARGALWECGCATLGAFOYAOAKPK
 VEELNVCNVSQFVSVNPKGYKVAFLPTDAGVATATGKTSATINHEWVGASLSY
 RLNSLPVYGVQWRSRATEDNIRIAQKPLPTAVLNLTAHNPSSLGNATLSTDSRS
 DPMQIVSQINKFKSRKAGVTGATLVADKWSLTAERLINERAAHVSGQFRP"
 54
 variation
 /gene="MOMP"
 /note="silent mutation"
 /replace="c"
 126
 variation
 /gene="MOMP"
 /note="silent mutation"
 /replace="g"
 320 a 254 c 246 g 350 t

BASE COUNT
 ORIGIN

Query Match 99.7%; Score 1098; DB 1; Length 1170;
 Best Local Similarity 100.0%; Pred. No. 1.1e-275;
 Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 4 ttgctgtagggaaccccttctgacgaagcttataatgattggtgacgaatggaaggt 63
 DB 70 TTGCCTGTAGGGAACCCCTTCTCATCCAGCTTATTAAATGATGGGTACAAATATGGAGGT 129
 QY 64 gtcgagagagatccttgagatccttgcctacttggctgagcagcgtattagcttgcgt 123
 DB 130 GGTGAGGAGATCCTTGGCGATCCTTGGCTACTTGGCTGCGAGCGCTATTAGCTTACGTCT 189
 QY 124 ggattttacggagactatgttttcgacgctatcttataaagttagatgcacataaacattt 183
 DB 190 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT 249
 QY 184 tctatgggagccaaactgactgagctgctgctgcaaaactataactactgctgagtagataga 243
 DB 250 TCTATGGGAGCCAAAGCTTACTGATCGCTGCTGCAAACTACTACTGCGGTAGATAGA 309
 QY 244 cctaaccggcctacaataagcatttacacgatgagcagcagcgtggttcactaagcagcctc 303
 DB 310 CCTAACCCGGCCTACAAATAGCATTTACAGATGACAGATGCTGTTTCAATATGACGGCTTC 369
 QY 304 attgcttaaacatttgggagcgttctgattgtttctgactttaggagccttcaatggt 363
 DB 370 ATTGCTTAAACATTTGGGATCGCTTGTATGTTTCTGCTACTTTAGGAGCTTCTTAATGTT 429
 QY 364 tacattaggaactctacagcgttcaatcctgttctgttattcggaggttaaggtact 423
 DB 430 TACATTAGAGAACTCTACAGCTTCAATCTCGTTGGTTTATTCCGAGTTAAAGGTACT 489
 QY 424 actgtaaatgcaatgaactacaaacgctttctttaaagtaacggaggtgttgaactttac 483
 DB 490 ACTGTAAATGCAATGAACACTCAACACGTTCTTTTAAAGTAACGGAGTGTGTGAACCTTAC 549
 QY 484 acagacacaccttcttctcttgaggcgttaggcgctctgagccttatgggaatgcggttgt 543

DB 550 ACAGACACTCTTTCTCTTGGAGCGTAGCGCTCTGTGGAGCTTATGGAAATGCGGTTGT 609
 QY 544 gcaacttgggagctgaattccaatatgcacagtcaccaacacctaagttgaagaactta 603
 DB 610 GCAACTTTGGGAGCTGAATTCCTCAATATGCACAGTCCAAACCTTAAAGTTGAAGA 669
 QY 604 gtgactctgaagctatcgaattctctgataaacaacccaagggcgtataaaggcgttct 663
 DB 670 GTGATCTGTACGTATCGCAATTCCTGTAAACAAACCAAGGGCTATAAAGGCGTTGCT 729
 QY 664 ttcccttggcacaacagcgtggtgctgacacagctactggaacaaagtctgcgaccatc 723
 DB 730 TTCCCTTGGCCAAACAGCGCTGGCGTAGCAACAGCTACTGGACAAAGCTTGGGACCATC 789
 QY 724 aattatcatgaatggcgaagtaggagcctctcttatctacagactaaactcttttagtcca 783
 DB 790 AATTATCATGAATGGCAAGTAGTAGGAGCTCTCTATCTTACAGACTAAACTCTTTAGTGCCA 849
 QY 784 tacattggagtacatgctcgcagcaacttttgcgtgctgataacatccgacttgcctcag 843
 DB 850 TACATTTGGAGTACAATGGTCTCGAGCAACTTTTGTATGCTGATAACATCCGCAATTCCTCAG 909
 QY 844 ccaaaactacctaagcgtgttttaaaacttaactgcatggaaccccttctttactagaaat 903
 DB 910 CCAAAACTTACCTACAGCTGTTTAAACTTAACTGCATGGAACCCCTTCTTTACTAGGAAAT 969
 QY 904 gccacagatgctcactactgattcgtctcagactcagactcagcaaatgtttcctgctcag 963
 DB 970 GCCACAGCATTTGCTTACTTACTTCTGATTCGATTCAGACTTTCATGCAAAATTTGTTCCCTGTCAG 1029
 QY 964 atcaacaagttaaactctagaaaagcttggagttgctgagcgtactgagcgtactttagtcat 1023
 DB 1030 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 1089
 QY 1024 gctgataaattggctacttactgagcagcagcgtctgttaataacagagagcgtcgcacgta 1083
 DB 1090 GCTGATAAATGGTCACTTACTTCTGACAGAGCTCGTTTAAATTAACGAGAGAGCTGCTCACGTA 1149
 QY 1084 tctgctcagttcagattc 1101
 DB 1150 TCTGCTCAGTTTCAGATTC 1167

RESULT 2
 AB033787
 ID AB033787 standard; DNA; PRO; 1170 BP.
 AC AB033787;
 SV AB033787.1
 DT 04-NOV-1999 (Rel. 61, Created)
 DT 04-NOV-1999 (Rel. 61, Last updated, Version 1)
 DE Chlamydomonas pneumoniae ompA gene for major outer membrane protein,
 DE complete cds.
 XX major outer membrane protein.
 XX Chlamydomonas pneumoniae
 OS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomonadales.
 RN 1-1170
 RP Hirakawa H., Shirai M.;
 RT Submitted (25-OCT-1999) to the EMBL/GenBank/DBJ databases.
 RL Hideki Hirakawa, Kyushu University, Graduate School of Genetic Resources
 RL Technology; 6-10-1 Hakozaki, Higashi-ku, Fukuoka 812-8581, Japan
 RL (E-mail:hirakawa@grt.kyushu-u.ac.jp, URL:http://www.grt.kyushu-u.ac.jp,
 RL Tel:81-92-642-3043, Fax:81-92-642-3043)
 XX

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic. - nucleic search, using sw model

Run on: February 7, 2002, 19:37:19 ; Search time 12230.8 Seconds
(without alignments)
1485.051 Million cell updates/sec

Title: US-09-391-606-14
Perfect score: 1101
Sequence: 1 agtgcctgttaggaaccc.....tatctggtcagtcagattc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo.hum.*
- 31: em.htgo.inv.*
- 32: em.htgo.rod.*
- 33: em.htg.hum.*
- 34: em.htg.inv.*
- 35: em.htg.rod.*
- 36: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1098	99.7	1170	1	AF131889	AF131889 Chlamydia
2	1098	99.7	1170	15	AB033787	AB033787 Chlamydia p
3	1098	99.7	1830	1	CHTOMMP	M69230 Chlamydia p
4	1098	99.7	20057	1	AE001652	AE001652 Chlamydia
5	1098	99.7	25150	1	AE002167	AE002167 Chlamydia
6	1098	99.7	300550	1	AP002547	AP002547 Chlamydia p
7	1096.4	99.6	1170	1	CHTOMPB	M64064 Chlamydia p
8	1086	96.8	1602	1	CPMOMPX	W72023 Chlamydia p
9	997.2	90.6	1170	1	CHTOMPEQ	L04982 Chlamydia p
10	967	87.8	999	1	CHTOMPAAI	M73038 Chlamydia p
11	668.4	60.7	670	1	AF131230	AF131230 Chlamydia
12	665.2	60.4	670	1	AF131229	AF131229 Chlamydia
13	606.6	55.1	1425	1	AF269259	AF269259 Chlamydia
14	600.2	54.5	1170	1	AF272945	AF272945 Chlamydia
15	600.2	54.5	1261	1	CHTOMPAAD	M73036 Chlamydia p
16	600.2	54.5	1337	1	CPOMP	X51859 Chlamydia p
17	600.2	54.5	1338	1	CHTOMPAAAA	L39020 Chlamydia p
18	600.2	54.5	1435	1	AF269256	AF269256 Chlamydia
19	595.8	54.1	1436	1	AF269267	AF269267 Chlamydia
20	592.6	53.8	1430	1	AF269266	AF269266 Chlamydia
21	590.6	53.6	1392	1	AF269279	AF269279 Chlamydia
22	585.8	53.2	1176	1	CHTOMPAAX	L25436 Chlamydia p
23	585.8	53.2	1412	1	AF269260	AF269260 Chlamydia
24	585.8	53.2	1421	1	AF269261	AF269261 Chlamydia
25	581.2	52.8	1482	1	AF269282	AF269282 Chlamydia
26	573.4	52.1	1200	1	AF269257	AF269257 Chlamydia
27	573.4	52.1	1200	1	AF269258	AF269258 Chlamydia
28	573.4	52.1	1227	1	CHTOMPAAH	M73037 Chlamydia p
29	573	52.0	1387	1	AF269280	AF269280 Chlamydia
30	571.8	51.9	1343	1	CPFPNMOMP	X61096 Chlamydia p
31	555.6	50.5	1434	1	AF269269	AF269269 Chlamydia
32	552.2	50.2	1107	1	CPDNAX	218756 Chlamydia p
33	543	49.3	1024	1	CHTOMPAAK	M73040 Chlamydia p
34	536	48.7	1058	1	CPAJ5617	AJ005617 Chlamydia
35	534.4	48.5	1058	1	CPAJ5613	AJ005613 Chlamydia
36	534.4	48.5	1058	1	CPAJ5614	AJ005614 Chlamydia
37	532.8	48.4	1058	1	CPAJ5615	AJ005615 Chlamydia
38	532.8	48.4	1058	1	CPAJ5618	AJ005618 Chlamydia
39	532.8	48.4	1058	1	CPJ004875	AF269265 Chlamydia
40	529.6	48.1	1445	1	AF269265	AF269265 Chlamydia
41	529.6	48.1	1610	1	CHTOMPPXX	L04980 Chlamydia p
42	528	48.0	1058	1	CPJ004874	AF269262 Chlamydia
43	528	48.0	1393	1	AF269262	AF269262 Chlamydia
44	528	48.0	1432	1	AF269264	AF269264 Chlamydia
45	528	48.0	1438	1	AF269263	AF269263 Chlamydia

ALIGNMENTS

RESULT 1

AF131889	AF131889	1170 bp	DNA	BCT	31-MAR-1999
LOCUS	Chlamydia pneumoniae major outer membrane protein (MOMP) gene,				
DEFINITION	complete cds.				
ACCESSION	AF131889				
VERSION	AF131889.1	GI:4545320			
KEYWORDS	Chlamydia pneumoniae				
SOURCE	Chlamydia pneumoniae				
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia				
REFERENCE	1 (bases 1 to 1170)				
AUTHORS	Sriram,S., Mitchell,W.M. and Stratton,C.W.				
TITLE	Multiple sclerosis associated with Chlamydia pneumoniae infection of the CNS				
JOURNAL	Neurology 50 (2), 571-572 (1998)				
MEDLINE	98145402				
REFERENCE	2 (bases 1 to 1170)				
AUTHORS	Mitchell,W.M., Tharp,A.C., Stratton,C.W. and Sriram,S.				
TITLE	Direct Submission				

GenCore: version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:07 ; Search time 96.2 Seconds
(without alignments)
366.620 Million cell updates/sec

Title: US-09-391-606-9
Perfect score: 2261
Sequence: 1 MWNPIGPGIDETERTPPAD.....SAAVVSAGVLPLOQVLWIRA 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues
Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_68.*
1: pir1.*
2: pir2.*
3: pir3.*
4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2214	97.9	715	2 I40729	hypothetical 76k p
2	2188	96.8	651	2 D72042	conserved hypothet
3	2188	96.8	651	2 E86581	CHLPN 76 kDa homol
4	451.5	20.0	647	2 G71490	hypothetical prote
5	201	8.9	973	2 C85693	probable membrane
6	188	8.3	1122	2 G64887	probable tail fibe
7	182.5	8.1	2055	2 T31110	extracellular matr
8	177.5	7.9	1822	2 S33441	EF protein - Strep
9	174.5	7.7	545	2 E84327	Htr7 transducer [i
10	174.5	7.7	545	2 T46811	halobacterial tran
11	171	7.6	627	2 F84194	Htr14 transducer [
12	166.5	7.4	1128	2 T30296	R27-2 protein - Tr
13	166.5	7.4	2232	2 T34434	hypothetical prote
14	160.5	7.1	544	2 T44938	transducer protein
15	158.5	7.0	1156	2 T34852	probable secreted
16	158	7.0	536	1 A47190	transducer protein
17	158	7.0	536	2 E84318	Htr1 transducer [i
18	158	7.0	810	2 T46810	halobacterial tran
19	158	7.0	810	2 F84327	Htr5 transducer [i
20	158	7.0	1365	2 T30822	lmp1 protein - Myc
21	156.5	6.9	641	2 C82206	methyl-accepting c
22	156	6.9	990	2 I51618	nucleolar phosphop
23	155	6.9	892	2 T50985	related to transcr
24	154.5	6.8	764	2 A84328	Htr2 transducer [i
25	154.5	6.8	765	1 T44946	transducer protein
26	154.5	6.8	1262	2 T22523	hypothetical prote
27	152.5	6.7	860	2 T14650	tail fiber protein
28	152.5	6.7	978	2 T14968	phage lambda-relat
29	152	6.7	1561	1 S06839	surface antigen sp

30 151.5 6.7 582 2 S24545 intermediate filam
31 151.5 6.7 1238 2 T03465 probable exonuclea
32 150.5 6.7 1528 2 A60338 surface antigen A
33 149 6.6 642 1 T44253 transducer protein
34 148 6.5 729 2 E70803 hypothetical prote
35 148 6.5 881 2 S56032 probable membrane
36 148 6.5 1556 2 A60988 saliva-interacting
37 147.5 6.5 1147 2 T35781 hypotheical prote
38 147.5 6.5 1302 1 JC6009 surface-located me
39 147 6.5 778 2 T48897 transducer protein
40 147 6.5 5327 2 T13564 microtubule-associ
41 146.5 6.5 446 2 H83098 hypothethical prote
42 146.5 6.5 571 2 D86164 hypothethical prote
43 146.5 6.5 643 2 H84305 Htr8 transducer [i
44 146 6.5 1714 1 S18644 multifunctional am
45 145.5 6.4 628 2 F84219 Htr16 transducer [

ALIGNMENTS

RESULT 1

I40729
hypothetical 76k protein - Chlamydothila pneumoniae (strain AR39)
C:Species: Chlamydothila pneumoniae, Chlamydia pneumoniae
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
C:Accession: I40729
R:Peréz-Melgosa, M.; Kuo, C.
Infect. Immun. 62, 880-886, 1994
A:Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76 kDa protein
A:Reference number: I40729; MUID:94156481
A:Accession: I40729
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-715 <RES>
A:Cross-references: GB:I23921; NID:g435961; PIDN:AAA23117.1; PID:g435962
A:Experimental source: strain AR-39
C:Comment: This is the hypothetical translation of a sequence that was reported as t

Query Match 97.9%; Score 2214; DB 2; Length 715;
Best Local Similarity 98.9%; Pred. No. 1.4e-110;
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;
Qy 1 MWNPIGPGIDETERTPPADLSAOGLEASAAKSAEAOIRIAGAEAKPKESKTSVERNSI 60
Db 257 LVNPIGPGIDETERTPPADLSAOGLEASAAKSAEAOIRIAGAEAKPKESKTSVERNSI 316
Qy 61 LRSVAVNMLSLADKLGSIASSNSSSTSRSDVDSTTATATPTPPPTSDDYKTQATAYDT 120
Db 317 LRSVAVNMLSLADKLGSIASSNSSSTSRSDVDSTTATATPTPPPTSDDYKTQATAYDT 376
Qy 121 IFTSTSLADIQAALVSLQDAVTNKDPTAATDEETAIAAEWETKNADATKVGQAITEALAKY 180
Db 377 IFTSTSLADIQAALVSLQDAVTNKDPTAATDEETAIAAEWETKNADATKVGQAITEALAKY 436
Qy 181 ASDNQAILDSLGKLTSPDLLQATALLQSVANNKAAELKEMQDNPPVPCKTPAIAQSLVD 240
Db 437 ASDNQAILDSLGKLTSPDLLQATALLQSVANNKAAELKEMQDNPPVPCKTPAIAQSLVD 496
Qy 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTOIAEAQ 300
Db 497 QTDATATQIEKDGNAIRDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTOIAEAQ 556
Qy 301 KKFPDPSPLQAEQOMVIOAEKDLNKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLL 360
Db 557 KKFPDPSPLQAEQOMVIOAEKDLNKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLL 616
Qy 361 DDAENETASILMSGFRQMIHMFNTENPDQAAQQLAAQARAARAAAGDDSAALADAAQK 420
Db 617 DDAENETASILMSGFRQMIHMFNTENPDQAAQQLAAQARAARAAAGDDSAALADAAQK 676
Qy 421 ALEALKAGKQOQOGLILNALGOIASAAVVSAGVLPLOQVL 459

Db 677 ALEAALGKAGQOQILNALGQIASAAVVSAGVLPQQVL 715
RESULT 2
D72042 conserved hypothetical protein CP0018 [imported] - Chlamydomonas reinhardtii (strain CW1)
N:Alternate names: chlpn 76 kda homolog_1 (ct622); hypothetical protein CP0018
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: D72042; D81623
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72042
A:Molecule type: DNA
A:Residues: 1-651 <ARN>
A:Cross-references: GB:AE001654; GB:AE001653; NID:g4377031; PID:g437703
A:Experimental source: strain CW1029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydomonas reinhardtii and Chlamydomonas reinhardtii AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: D81623
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <REA>
A:Cross-references: GB:AE002161; NID:g7188948; PID:AAF37914.1; PID:g718894
A:Experimental source: strain AR39, HL cells
C:Comment: This sequence was originally identified as homologous to part of a sequence
PIR:H1490).
C:Genetics:
A:Gene: CP0018

Query Match 96.8%; Score 2188; DB 2; Length 651;
Best Local Similarity 98.9%; Pred. No. 3e-109;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MYPNIGPGPIDETERTPPADLSAQGLEASAAKSAEQAQIAGAEAKPKESKTSVERWSI 60
Db 1 MYPNIGPGPIDETERTPPADLSAQGLEASAAKSAEQAQIAGAEAKPKESKTSVERWSI 60
QY 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120
Db 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120
QY 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180
Db 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180
QY 181 ASDNQAILDSLGKLTFFDLLOALQSVAANNKAAELKEMQDNPVPGKTPAIQAQSLVD 240
Db 181 ASDNQAILDSLGKLTFFDLLOALQSVAANNKAAELKEMQDNPVPGKTPAIQAQSLVD 240
QY 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAQ 300
Db 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAQ 300
QY 301 KKFPDPSPILOEAQWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360
Db 301 KKFPDPSPILOEAQWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360
QY 361 DDAENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAGDSDSAAALADAQK 420
Db 361 DDAENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAGDSDSAAALADAQK 420
QY 421 ALEAALGKAGQOQILNALGQIASAAVVSAGVLP 454
Db 421 ALEAALGKAGQOQILNALGQIASAAVVSAGVLP 454

RESULT 4
G71490 hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
N:Alternate names: chlpn 76kda homolog CT622
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: G71490
R:Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitc
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia
A:Reference number: A71570; MUID:95000809
A:Accession: G71490
A:Molecule type: DNA
A:Residues: 1-647 <ARN>
A:Cross-references: GB:AE001333; NID:g3329068; PIDN:AAC68226.1; PID:g3

Db 677 ALEAALGKAGQOQILNALGQIASAAVVSAGVLPQQVL 715
RESULT 2
D72042 conserved hypothetical protein CP0018 [imported] - Chlamydomonas reinhardtii (strain CW1)
N:Alternate names: chlpn 76 kda homolog_1 (ct622); hypothetical protein CP0018
C:Species: Chlamydomonas reinhardtii, Chlamydomonas reinhardtii
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: D72042; D81623
R:Kaiman, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydomonas reinhardtii and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72042
A:Molecule type: DNA
A:Residues: 1-651 <ARN>
A:Cross-references: GB:AE001654; GB:AE001653; NID:g4377031; PID:g437703
A:Experimental source: strain CW1029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydomonas reinhardtii and Chlamydomonas reinhardtii AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: D81623
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <REA>
A:Cross-references: GB:AE002161; NID:g7188948; PID:AAF37914.1; PID:g718894
A:Experimental source: strain AR39, HL cells
C:Comment: This sequence was originally identified as homologous to part of a sequence
PIR:H1490).
C:Genetics:
A:Gene: CP0018

Query Match 96.8%; Score 2188; DB 2; Length 651;
Best Local Similarity 98.9%; Pred. No. 3e-109;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MYPNIGPGPIDETERTPPADLSAQGLEASAAKSAEQAQIAGAEAKPKESKTSVERWSI 60
Db 1 MYPNIGPGPIDETERTPPADLSAQGLEASAAKSAEQAQIAGAEAKPKESKTSVERWSI 60
QY 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120
Db 61 LRSVAVNLSADKLGIASSNSSTSRSDVSTTATPTPPPTSDDYKTAQATAYDT 120
QY 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180
Db 121 IFTSTSLADIQAALVSLQDAVTNKTATDETAIAAEWETKNADAIVKGAQITELAKY 180
QY 181 ASDNQAILDSLGKLTFFDLLOALQSVAANNKAAELKEMQDNPVPGKTPAIQAQSLVD 240
Db 181 ASDNQAILDSLGKLTFFDLLOALQSVAANNKAAELKEMQDNPVPGKTPAIQAQSLVD 240
QY 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAQ 300
Db 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAQ 300
QY 301 KKFPDPSPILOEAQWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360
Db 301 KKFPDPSPILOEAQWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360
QY 361 DDAENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAGDSDSAAALADAQK 420
Db 361 DDAENETASTILMSGFQMIHMFNTENPDQAQOELAAQAAKAGDSDSAAALADAQK 420
QY 421 ALEAALGKAGQOQILNALGQIASAAVVSAGVLP 454
Db 421 ALEAALGKAGQOQILNALGQIASAAVVSAGVLP 454

A:Experimental source: serotype D, strain UW-3/Cx
C:Comment: This sequence was originally identified as homologous to part of a sequence
PIR:E72042).
C:Genetics:
A:Gene: CT622

Query Match 20.0%; Score 451.5; DB 2: Length 647;
Best Local Similarity 29.6%; Pred. No. 6.7e-17;
Matches 138; Conservative 85; Mismatches 192; Indels 51; Gaps 14;
QY 2 VNPVIGPCTDETERPPADLSAQGLEASAAKSAEQAORAGAEAKPKESKTSVERWSIL 61
DB 15 MPPIINGQI-----ASNSETKESTKESEA-----SPSASSVSSWSFL 52
QY 62 RSVAVNALMSLADKLGFIASSNSSTSRSDVDTTATPTPPPTDDY---KTOAQTV 118
DB 53 SAKHALISLRD--ALLNKNSSPTDSL-OLEASTSTSTVTRVARDYNEAKSNFDTAK 108
QY 119 DTFTSTSLADIAQALVSLQDAVTNKTDAATDEETAIAAEWETKNADAIVGAQITELA 178
DB 109 SGLNATTIAEYETKMAQLMADLMERLAKQAEVTRIKALQEKQEVID---KLNQLV 165
QY 179 KYASDNOAILSLGKLTSLDLOLQVAVNNKAAELLKEMQDNPVPGTKPTAQSL 238
DB 166 KLEKQNTLKTETTTTDSADQIPAINSOLEINKNSADQIKDEGQNI---SYEAVLTNA 222
QY 239 VQOTDATATQIEKDNAGDAIFAGQVAGNAGVENAKNSNISNIDSKAAIATAKTOIAE 298
DB 223 GEVIRASSPAGIKLQALQISVDAGQSOAVLQAOQNNSPDIAATKLIIDAAETKVE 282
QY 299 AQKK---FPDSPILQAEQVMQVIAEKDKLNKIPADGSDVPNPPTTVGGSKQOQSGSIS- 354
DB 283 LKQHTGLTSLPLVKAEEQISOAKDIQEIKPS-GSDIPVIGPS-GSAASAGSAGVALK 340
QY 355 -----RVSMILLDDAENETASILMSGFROMIHMFNTENPDQSOAQOELAAQARA- 405
DB 341 SSNNSGRIILLDDVDNEMAAATAMQGFMSIEQFNVPNNPATAKELOAMEAQLTMSDQLV 400
QY 406 AGDSDAAAALADAQALEAKAGQOQOQILNALGQIASAAVVSAG 451
DB 401 GADGELPAEIQAIKDALAQAL-KQPTDGLATAMQGVAAAKVGG 445

RESULT 5
C85693
probable membrane protein of prophage Cp-933X Z1918 [imported] - Escherichia coli (strain
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 31-Mar-2001
C:Accession: C85693
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouisis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: C85693
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-973 <STO>
A:Cross-references: GB:AE005174; NID:g12514847; PIDN:AAG56007.1; GSPDB:GN00145; UWCP:Z19
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z1918

Query Match 8.9%; Score 201; DB 2: Length 973;
Best Local Similarity 22.9%; Pred. No. 0.0024;
Matches 120; Conservative 64; Mismatches 210; Indels 130; Gaps 14;
QY 26 LEASAAKSAEQAORAGAEAKPKESKTSVERWSILRSVAVNALMSLADKLGFIASSNSSSS 85
DB 112 VEEVARNASAVAQNTAAAKKSADASTSAREATHATDAADSARAASTSAGQAAS-SAQS 170

QY 86 TSSADVDTTATPTPPPTSDDYKTOAQATYDTFTSTSLADIAQALVSLQDAVTNIK 145
DB 171 ASSSACTASTKATEASKSAASAAESSKSAATSAAGAKTSET-----NAAVSQOASATSAS 225
QY 146 DTAATDEETAIAAEWETKNADAIVGAQITELAKYASDNOAILDLSLGLTKLTSLDLOLQALL 205
DB 226 TATTKASEAASSA---RDASASKEAAKSETSAASSASAAASSASATAAGNSAKAAKTS-- 279
QY 206 QSVANNKAAELLKEMQDNPVPGTKPTA-----IAQSLVDQDTATATQIEK- 252
DB 280 ---ETNAKSETAEQASAAAGSKTAALASASAASSTAGQASASATAAGKSAESAASSA 336
QY 253 -----GNAIGDAYFAGQNASGA---VENAKSNNSIS 280
DB 337 STATTKAGEATEQASAAASAAKSETNAKASETSAESKKTAAASSASASSASASSAS 396
QY 281 -----NIDSKAAIATAKTOIAE-----AOKKFPDSPILQAEQVMQVIAEKDKL 324
DB 397 ASKDEATROASAAKSSATTASTKATEAGSATAAAGSKSTAESAATRAETAARAEADIAS 456
QY 325 NIKPADGSDVPNPPTTVGGSKQOQSGSISIRVSMILLDDAENETASILMSG----- 374
DB 457 AVALEDAS-----TTKKGIVQLSSATNTSTESLAATPRAVKAAYELANGKYTAQDATTA 510
QY 375 FROMIHMFNTENPDQSOAQOELAAQARAAGD-----D 409
DB 511 QKGIVQLSSATNTSTSEM---LAATPKSVKAAVYDLANGKYTAQDATTAQKGIQLSSATN 566
QY 410 SAAALADAKQALEAKAGQOQOQILNALGQIASAAVVSAGVL 453
DB 567 SASSETLAATPAKVAKAAND-----NANGRVPSARKVNGKAL 601

RESULT 6
G64887
probable tail fiber protein gp37 - Escherichia coli
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 21-Jul-2000
C:Accession: G64887; T09189
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A:Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: G64887
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-1122 <BLAT>
A:Cross-references: GB:AE0000234; GB:U000096; NID:g1787633; PIDN:AAC74454.1; PID:g17876
A:Experimental source: strain K-12, substrain MG1655
R:Albal, H.; Baba, T.; Fujita, K.; Hayashi, K.; Inada, T.; Isono, K.; Itoh, T.; Kasai,
M.; Motomura, K.; Nakade, S.; Nakamura, Y.; Nashimoto, H.; Nishio, Y.; Oshima, T.; Sai
moto, Y.; Horiuchi, T.
DNA Res. 3, 363-377, 1996
A:Title: A 570-Kb DNA sequence of the Escherichia coli K-12 genome corresponding to t
A:Reference number: Z16603; MUID:97251357
A:Accession: T09189
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 3-1122 <ATB>
A:Cross-references: GB:AE0000234; GB:U000096; NID:g1787633; PIDN:AAC74454.1; PID:g17876

Query Match 8.3%; Score 188; DB 2: Length 1122;
Best Local Similarity 23.0%; Pred. No. 0.014;
Matches 103; Conservative 68; Mismatches 203; Indels 74; Gaps 13;
QY 8 GPIDETERTPPA---DLQAQGLEASAAKSAEQAORAGAEAKPKESKTSVERWSILRS 63
DB 93 GAMTEDDARPEALRRFELM---VEEVARNASAVAQNTAAAKKSADASTSAREATHAAD 149
QY 64 AVNALMSLADKLGFIASSNSSTSRSDVDTTATPTPPPTSDDYKTOAQATYDTFT 123

Db 150 AADSAARASTSAGQAAS--SAQSASSAGTASTKATEASKSAAAAESSKAAATSAKAAT 208
QY 124 STSLADIOALYSLQDAVNIKDTATDEETALAEWETKNADAIVKVGQITELAKYASD 183
Db 209 SETNAS-----ASLQSAATASATTTKASEAATSA-----RDAASKEAAKSETNASSA 259
QY 184 NOAILDSLGLTDFDLQTLALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTD 243
Db 260 SSAASATAAGNSAKAAT'S-----ETNARSETAAGQSASAAAGSKTAAASASASAT'S 314
QY 244 ATATQTEKQGNAGIDGAYFAGQNASGAVENAKSNNSISNIDSAKAAIATATAKTQIAEAKKF 303
Db 315 A-----GQASASATAAGKSA-----ESAASSASTATTKAGEATEQ- 349
QY 304 PDSPILOEAFQVIOAEKDLKNIKPADGSDVPNPGTTVGGSKQOQSGSIRVSMMLDDA 363
Db 350 -----ASAARSAASAKTSETNAKASETSAESKTTAAASASASASASASK--DEA 402
QY 364 ENETASILMSGFROMIHMFNTENPDSSQAQOELAAQA-----RAAKAAGDSDSAAALAD 417
Db 403 TRQASAKSSATTASTKATEAGSATAAQAQSKSTAESAATRAETAAKRAEDIASAVALD 462
QY 418 AQKALEALGKAGQOQOGLNALGQIASA 445
Db 463 ASTT-----KRGIV-----QLSSA 476

RESULT 7
T31110
extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C:Species: Abiotrophia defectiva
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31110
R:Manganelli, R.; van de Rij, I.
Infect. Immun. 67, 50-56, 1999
A:Title: Cloning and characterization of emb, a gene encoding the major adhesin of Strep-
A:Reference number: Z20988; MUID:99081722
A:Accession: T31110
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-2055 <MAN>
A:Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AAD03320.1
C:Genetics:
A:Gene: emb

Query Match 8.1%; Score 182.5; DB 2; Length 2055;
Best Local Similarity 22.9%; Pred. No. 0.06;
Matches 109; Conservative 76; Mismatches 189; Indels 101; Gaps 21;
QY 19 ADLSAQGLEASAAKSAEQRIAGAA-----KPKESKTSVERWSILRSVAVNALMSLD 73
Db 1371 ATAKNAIDAATSNDDKATQNEGTQAINAVPQTTPRAKTD-----AKNAVQAAAD 1420
QY 74 KLGIASSNSSST-----SRSDVDSTTATPTPPPTSDDYKTQATAYDTITSTSLA 128
Db 1421 RKXDAIENDPNLTREEKVAAKAVDA-----EAKKAKDAIDAATNSA 1462
QY 129 DIQAALYSLQDAVNIKDTATDEETALAEWETKNADAIVKVGQITELAKYASDQNA 186
Db 1463 DVTAKQNEGTKAINDVPQTPTAKTDNAKNAVQAAADAKK--DAIEKDPNLTREEKDA--KA 1519
QY 187 ILDSLGLTDFDLQTLALQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTD 244
Db 1520 KVDAEAKK--DAIDAATSNADVTAKQNEGTKAINDVPQTPT--AKTDA--KNAVQAAAD 1575
QY 245 TATQIEKQGNAGIDGAYF-----GQNASGAVENAKSN-----NSISNI 282
Db 1576 KKAIEKDPNLTREEKDAAKAVDAEAKKAKDAIDAATSNADVTAAQKDGAKNAINAVPQT 1635
QY 283 DSAKAAIATATAKTQIAEAKK--KFPDSPILOEAFQVIOAEKDLKNIKPADGSDVPNPGTTV 341
Db 1636 PTAKTDKNAVQAAADKDAIENDANLTREEKDAAKAVDAEATKAKNAIDAATSNADV 1695

QY 342 GGSKOQGS--SIGSI--RVSMMLDDAEN---ETASILMSGFROMIHMFNTEN-----PDSQ 390
Db 1696 TAKQNEGTKAINDVPQTPTAKTDNAKNAVQAAATDKKSAI-----ENDPALTREEKD 1746
QY 391 AAQOELAAQARAAGAGDSDSAAALADAAQKALEAALGKAGQOQOGLNALGQIASA 445
Db 1747 AAKAKVDAEATKAKNAIDAATSNADVTAAQK-----DAG--KNAINAVPQTPTA 1792

RESULT 8
S33441
EF protein - Streptococcus suis
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are abs-
A:Reference number: S33441
A:Accession: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SMI>
A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

Query Match 7.9%; Score 177.5; DB 2; Length 1822;
Best Local Similarity 23.6%; Pred. No. 0.095;
Matches 116; Conservative 77; Mismatches 195; Indels 103; Gaps 21;
QY 13 TERTPPADLSAQGLEASAAKSAEQRIAGAAKAPRES--KTDSEVERWSIL--RSAVNALMS 70
Db 1201 TEKAKAELAGEKSLDTGCKEARDAVELAKDELAKELAEIRTEEBEATKIVEKLAEDTRKA 1260
QY 71 LADKLGIASSNSSSTSRGAD--VDSTTATPTPPPTSDDYKTQATAYDTITSTSLAD 129
Db 1261 IEDNPNLSDEDKQAEIKKLTDAVAKTLATI-----RDNAKRTQEAKEA-----QALAD 1309
QY 130 IQAALYSLQDAVNIKDTATDEETALAEWETKNADAIVKVGQITELAKYASDNOAI 187
Db 1310 LEKA-----KETQKIADKKAIDRLTILVKDGELEATKQDAKNAKIAKAAKAEIASNPN 1364
QY 188 LDSLGLTDFDLQTLALQSVANNK-----AELLKEMQDNVPVPGKTPAIAQSLVD 240
Db 1365 LTDAEKKTFTD-----AVDAEVAKANDAISAATSPADVQKE--EDAGVA-----AIAEDVLD 1414
QY 241 OTDATA--TQIEKQGNAGIDGAYFAGQNASG-----VENAKSNNSISNIDSAKAAI 289
Db 1415 AAKODAKNKIAKDAKAAKAEAGISNPNLTDAEKKTTFTDAVDAEVAKANDAISA----- 1465
QY 290 ATAKTQIAEAKKFPDSPILOEAFQVIOAEKDLKNIKPADGSDV-----PNPGTT--- 340
Db 1466 --AATSPADVQKE--EDAGVAAIAEDVLDAAKQDAKN--KIAKESDAKSAIDANPNLTDAE 1521
QY 341 -----VGGSKQOQSGSIRVSM-----LDDAENETASILM 372
Db 1522 KESAKKAVDAKAAATDAIDASTSPVEAQSAEDKGVGSIADQVLDAAKQAKNKIAKEVA 1581
QY 373 SGFQMIHMFNTENPDSSQAQOELAAQARAAGDSDSAAALADAAQKALEAALGKAGQO 432
Db 1582 AAKEAIDANPNLTDAEKEASKAVDADAKATTDAID--ASTSPVEAQSAEDKGVGSI--R 1637
QY 433 QGILNALGQIA 443
Db 1638 QDVLDRAKQDA 1648

RESULT 9
E84327
Htr7 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001

C:Accession: E84327
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon
Jung, K.H.; Alam, M.; Freitas T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jablon
A:Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: E84327
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-545 <STO>
A:Cross-references: GB:AF004437; NID:g10581214; PIDN:AAG19985.1; GSPDB:GN00138
C:Genetics:
A:Gene: htr1
C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 7.7%; Score 174.5; DB 2; Length 545;
Best Local Similarity 21.7%; Pred. No. 0.029;
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;

Qy 12 ETERTPPADLSAOGLEASAAKSAEAGRIAGAEAKPKESKTSVERWILRSVAVNALMSL 71
Db 118 EEERA-----EAERAREAKQKQAEAEER---QTAEASAKQDARERSAEIEQLAADLESQ 169
Qy 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPT-SDYKTAQATYDTI-----FTST 125
Db 170 ATEVG-ATLEASDGLTARVDATTDNAEIAEAVATVNDMLTTWERTIDEIOGFSTNVTT 228
Qy 126 SLADIQAALVSLQDAVTNIKT-----AATDEE-----TAIAAEWTKNADAIKVGQAITE 176
Db 229 ASREATAGAKETQDASQTSVSEVQIEAAGTDQREQLSEAEEMDSYATVEEVAATAQS 288
Qy 177 LAKYASDNOAILDSLGKLT-----SFDLLQTLALQSVAANNKAAELKEMODNPVPGK 230
Db 289 VADTAADTTDVA-TAGKQTAEDAIDAIVQETMQTTVANVDALEDLLEIDT----- 340
Qy 231 TPATQAQSLVQDTAT-----ATQIEKDGNAIGD--AYFAGQNASGAVENAKSNNS 278
Db 341 ---IAELISDAEQTNMLALNANIEAARAGSGGSGDGFVAVDEVKELATESQSAKD 397
Qy 279 ISN-IDSAAKAIATAKTOIAEAKKFPDPSILOEAEQMVIAEKDLKNIKPADGSDVNP 337
Db 398 IAELEEVQSTATTVEI-----RVAEQRVNCGAAVEETVDAFGAVTENI 444
Qy 338 GTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMSGFRQMIHMFNTENPDPSQAAQBELA 397
Db 445 QETTDGVQE-----ISQAMDEQAQSRVSVSS-----VDDIATISQATA 483
Qy 398 AQARAAKAGDSDRAALADAKALEAALGKAGQOQGILN 437
Db 484 DRAENVSAASEEQ-TASITEVTSLSQSLAAQAQDTLEDRLN 522

RESULT 10
T46811
halobacterial transducer protein V [imported] - Halobacterium salinarum
C:Species: Halobacterium salinarum
C:Date: 17-Mar-2000 #sequence_revision 17-Mar-2000 #text_change 15-Sep-2000
C:Accession: T46811
R:Rudolph, J.; Nordmann, B.; Storch, K.F.; Gruenberg, H.; Rodewald, K.; Oesterhelt, D.
FEMS Microbiol. Lett. 139, 161-168, 1996
A:Title: A family of halobacterial transducer proteins.
A:Reference number: 224094; MUID:96275896
A:Accession: T46811
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-545 <RUD>
A:Cross-references: EMBL:X95589; NID:g1435130; PIDN:CAA64842.1; PID:g1435132
C:Genetics:
A:Gene: htp
C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 7.7%; Score 174.5; DB 2; Length 545;
Best Local Similarity 21.7%; Pred. No. 0.029;
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;

Qy 12 ETERTPPADLSAOGLEASAAKSAEAGRIAGAEAKPKESKTSVERWILRSVAVNALMSL 71
Db 118 EEERA-----EAERAREAKQKQAEAEER---QTAEASAKQDARERSAEIEQLAADLESQ 169
Qy 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPT-SDYKTAQATYDTI-----FTST 125
Db 170 ATEVG-ATLEASDGLTARVDATTDNAEIAEAVATVNDMLTTWERTIDEIOGFSTNVTT 228
Qy 126 SLADIQAALVSLQDAVTNIKT-----AATDEE-----TAIAAEWTKNADAIKVGQAITE 176
Db 229 ASREATAGAKETQDASQTSVSEVQIEAAGTDQREQLSEAEEMDSYATVEEVAATAQS 288
Qy 177 LAKYASDNOAILDSLGKLT-----SFDLLQTLALQSVAANNKAAELKEMODNPVPGK 230
Db 289 VADTAADTTDVA-TAGKQTAEDAIDAIVQETMQTTVANVDALEDLLEIDT----- 340
Qy 231 TPATQAQSLVQDTAT-----ATQIEKDGNAIGD--AYFAGQNASGAVENAKSNNS 278
Db 341 ---IAELISDAEQTNMLALNANIEAARAGSGGSGDGFVAVDEVKELATESQSAKD 397
Qy 279 ISN-IDSAAKAIATAKTOIAEAKKFPDPSILOEAEQMVIAEKDLKNIKPADGSDVNP 337
Db 398 IAELEEVQSTATTVEI-----RVAEQRVNCGAAVEETVDAFGAVTENI 444
Qy 338 GTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMSGFRQMIHMFNTENPDPSQAAQBELA 397
Db 445 QETTDGVQE-----ISQAMDEQAQSRVSVSS-----VDDIATISQATA 483
Qy 398 AQARAAKAGDSDRAALADAKALEAALGKAGQOQGILN 437
Db 484 DRAENVSAASEEQ-TASITEVTSLSQSLAAQAQDTLEDRLN 522

RESULT 11

F84194
Htr14 transducer [imported] - Halobacterium sp. NRC-1
C:Species: Halobacterium sp. NRC-1
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 16-Feb-2001
C:Accession: F84194
R:Ng, W.V.; Kennedy, S.P.; Mahairas, G.G.; Berquist, B.; Pan, M.; Shukla, H.D.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jung, K.H.; Alam, M.; Freitas T.
Proc. Natl. Acad. Sci. U.S.A. 97, 12176-12181, 2000
A:Authors: Hou, S.; Daniels, C.J.; Dennis, P.P.; Omer, A.D.; Ebhardt, H.; Lowe, T.M.; Lasky, S.; Leithauser, B.; Keller, K.; Cruz, R.; Danson, M.J.; Hough, D.W.; Maddocks, D.G.; Jung, K.H.; Alam, M.; Freitas T.
A:Title: Genome sequence of *Halobacterium* species NRC-1.
A:Reference number: A84160; MUID:20504483
A:Accession: F84194
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-627 <STO>
A:Cross-references: GB:AF004437; NID:g10579976; PIDN:AAG18922.1; GSPDB:GN00138
C:Genetics:
A:Gene: htr14
C:Superfamily: Halobacterium salinarum transducer protein htr1

Query Match 7.6%; Score 171; DB 2; Length 627;
Best Local Similarity 23.4%; Pred. No. 0.053;
Matches 120; Conservative 87; Mismatches 184; Indels 122; Gaps 27;

Qy 17 PPADLSAOGLEASAAKSAEAGRIAGAEAKPK-----ESKTSVERWS-ILRSVAVNALMSL 71
Db 145 PALDESVPAGGESITEMADSLEAYTAEDKRTAEHQAELEKRSQALVDALSEA 204
Qy 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDYKTAQATYDTI-----OTAVDTIFTS 124
Db 205 TD-----AARAGDL--TATVDAAALDVTDDHRAVEDFNQLETIADTISDI 249

QY 125 TSLADIQAAL-----VSLQDAVTNKTDA--ATDEETAIAAEWETKNADA 167
Db 250 QSFDAVLAVSRITTDERRVDAVADRSAAVSESVTEIADGANQOITNQLNIAAEMDTVSATV 309
QY 168 IKVGAQITELAKYASDQAIIL-----DSLGLK-TSFDLLQTLALLOSVAANNKAAELLKEMQ 222
Db 310 EEIAASANDVAKTA---QAADRDGEGEVEETIEALRALREQSQAETVESLAAEVE 366
QY 223 DNPVPGKTPAIAQSLVDOTDATA--TQIE-----KQNAIGDAYFAGONASGAVENAKS 275
Db 367 R---IDGIT-ALIEDIAEETNNLALNASIEAARTGSDGD--GFVAVADDEVKDLAEETREQ 420
QY 276 NNSISNIDSA-----KAAIATATQIAEAKQKPPDSP-ILQEAQMVVIOAEKDLKNIKP 328
Db 421 AADISEIVDAYTEKAEDASIAIGVD-AEVERKITKAEGLVRDFEAV-----DEVANVNH 475
QY 329 A--DGSVDVNPNG-----TTVGGSKQOOGSSI-----GSRVSMILLDD 362
Db 476 AVEISDATDQGAQSVTDVVGMEVASVSEETAESDTVADNAAEQTDATDEVADQMD 535
QY 363 AENETAST--LMSCFROMIHMFENTENPDSSQAQOELAAQARA--AAGDDSSAAALADA 418
Db 536 LAETUAAALAGMLDDF-----TVPADAGTADQOSVADDSPTAQPPAADDEPAAAVVDQP 587
QY 419 QKALEAALGRAGQOQOGLNALGQIASAAVVSAG 451
Db 588 QPASDAE-----DEGVDPDSGGE--SVAVSDGG 613

RESULT 12
T30296
R27-2 protein - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C:Date: 22-Oct-1999 #sequence_revision 22-Oct-1999 #text_change 17-Mar-2000
C:Accession: T30296
R:Otsu, K.; Donelson, J.E.; Kirchhoff, L.V.
Mol. Biochem. Parasitol. 57, 317-330, 1993
A:Title: Interruption of a Trypanosoma cruzi gene encoding a protein containing 14-amin
A:Reference number: 220813; MUID:93165082
A:Accession: T30296
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1128 <GT>
A:Cross-references: EMBL:L04603; NID:g385171; PID:g1256742; PIDN:AAA96494.1
C:Superfamily: neurofilament triplet H protein

Query Match 7.4%; Score 166.5; DB:2; Length 1128;
Best Local Similarity 23.9%; Pred. No. 0.2;
Matches 110; Conservative 71; Mismatches 202; Indels 77; Gaps 19;

QY 10 IDERTPPADLS---AOGLEASAANKSAEAOQRIAGAE--KPKESKTSVERWSILRSA 64
Db 642 VAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAE 701
QY 65 VNALMSLADKLIGIASNSSSSSTSRADVDSTTATPTPPPTSDDYKTQATYDTIFTS 124
Db 702 EKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEAT--KVAEAEKQKAAE-----TK 753
QY 125 TSLADIQAALVSLQDAVTNKTDAATDEETAIAAEWETKNADA 177
Db 754 VAEAEKQKAAEATKVAEAE--KQKAA--EATKVAEAEKQKAAEATKVAEAEKQKAAEATKV 810
QY 178 AKYASDQNALDLSGLKTSFDLLQTLALLOS--ANNKAAELLK-----EMQDNPVVPCKT 231
Db 811 AEAEKQKAA--EATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVA 867
QY 232 PAIAQSLVDOTDATAPTQIEKDGNAIGDAYFAGONASGAVENAKSN-----NSISIDS 284
Db 868 EAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEK 927
QY 285 AKAAIATAKQIAEAAQK-KFPDSPILQEAQ-----VIOAEKDLKNIKPADGSDVPN 336

Db 928 QKAAEA---TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQ-----KAAEATKVAE 980
QY 337 PCTTVGSGKQGGSIGIRVSMILLDDAENETASILMSGFROMIHMFENTENPDSSQAQOEL 396
Db 981 -----EAKQKAAEATKVA-----EAKQKAAEATK-----VAEAEKQKAAEATK 1019
QY 397 AAQARAQAAGDDSSAAALADAQKALEAALGKAGQOQOGL 436
Db 1020 VAEAEKQKAA--EATKVAEAEKQKAAEAAKAMESQKORFL 1057

RESULT 13
T34434
hypothetical protein K06A9.1a - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 18-Feb-2000
C:Accession: T34434
R:Geisel, C.; Gattung, S.
submitted to the EMBL Data Library, December 1996
A:Description: The sequence of C. elegans cosmid K06A9.
A:Reference number: Z21525
A:Accession: T34434
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-2232 <GEI>
A:Cross-references: EMBL:U80846; PIDN:AACT0890.1; GSPDB:GN00028; CESP:K06A9.1a
A:Experimental source: strain Bristol N2; clone K06A9
C:Genetics:
A:Gene: CESP:K06A9.1a
A:Map position: X
A:Introns: 38/1; 75/3; 103/3; 132/2; 158/2; 222/1; 1088/1; 1367/1; 2039/1; 2049/1; 21;

Query Match 7.4%; Score 166.5; DB:2; Length 2232;
Best Local Similarity 19.7%; Pred. No. 0.47;
Matches 79; Conservative 70; Mismatches 174; Indels 79; Gaps 11;

QY 4 PIGPGPIDETERTPPADLSAQ---GLEASAANKSAEAOQRIAGAE--KPKESKTSVERWS 59
Db 598 PSSOSPAPNTGSTTPSTQSSPSPMNPSSSTPTGSSQSTITPEGSTASSPTGSTGTF 657
QY 60 ILRSVAVNALMSL--ADKLIGIASNSSSSSTSRADVDSTTATPTPPPTSDDYKTQATA 117
Db 658 SVATEVTSQSVNPVSGSLTQSTNSPSPSLSPSTSCMSLTSEPPSS---TQSSGA 713
QY 118 YDTIFT--STSLADIQAALVSLQDAVTNKTDAATDEETAIAAEWETKNADA 176
Db 714 QSTLTTPSPNPQSSTLSSTSGATTSAGTT-----MTSPSSSSVSG----- 759
QY 177 LAKYASDQNALDLSGLKTSFDLLQTLALLOSVAANNKAAELLKEMQDNPVVPCKTPTAIAQ 236
Db 760 -SSOGSTSPRASTTSGEMTSQGSTQT-----PGSSVSTSA 793
QY 237 SLVDQTDATATQIEKDGNAIGDAYFAGONASGA---VENAKSNNSISNIDSAKAAIATAK 293
Db 794 AILTSTQSVSTNSPSTGSTRPSTVSGSTSGSTVTVGSTEASTSGSVASSSPAPSTSQ 853
QY 294 TQTAEAKKFPDPSILOEAQMVVIOAEKDLKNIKPADGSDVNPNG-----TTVGGSKQ 347
Db 854 N-----PNPSTSGSSMITQSPYPSQSTSPVESSTTTPSPGPTTLTSTSPSPSQ 903
QY 348 GSSIGTSIRVSMILLDDAENETASILMSGFROMIHMFENTENPD 389
Db 904 STTIGS-----TCGSTSPGISTTSEMTSQGSTQTGCS 936

RESULT 14
T44938
transducer protein htpv [similarity] - Halobacterium salinarum
N:Alternate names: methyl-accepting taxis protein htl; transducer protein htl; trans
C:Species: Halobacterium salinarum
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 02-Sep-2000

C:Accession: T44938
R:Zhang, W.; Brooun, A.; McCandless, J.; Banda, P.; Alam, M.
Proc. Natl. Acad. Sci. U.S.A. 93, 4649-4654, 1996
A:Title: Signal transduction in the archaeon Halobacterium salinarum is processed through a transmembrane protein
A:Reference number: 222804; MUD:96209786

A:Accession: T44938
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-544 <2HA>
A:Cross-references: EMBL:U53365; NID:gl589744; PIDN:AAB17519.1; PID:gl621047
A:Note: the source is designated as Halobacterium salinarum
C:Genetics:
A:Gene: htpV; htp5; htrVII; htrI
C:Superfamily: Halobacterium salinarum transducer protein htrI
C:Keywords: methylated amino acid; signal transduction; transmembrane protein
F:246-502/Region: MCP signalling domain similarity

Query Match 7.1%; Score 160.5; DB 2; Length 544;
Best Local Similarity 22.3%; Pred. No. 0.16;
Matches 101; Conservative 74; Mismatches 183; Indels 95; Gaps 19;

QY 12 ETERPPADLSAAGLEASAAKSAEARIAGAEAKPKESKTDSDVERWILRSVAVNALMSL 71
DB 118 EEERA-----EERAREAKQKQAEER-OTAEASQGTGRTQRE---IEOLAADLESQ 168
QY 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPT--SDDYKTAQAT-----AYDTIFTS 124
DB 169 ATEVG-ATLEASDGLTARVDATTDNAEIAEVATVVDMLTMTERTIDEIQGFSTNVT 227
QY 125 TS-----LADQALVSLQDAVNTIKDTAATDEE---TAIAEWETKNADAIVKGAQI 174
DB 228 ASREATGAKETIQVASQTVSESVQEI--AAGTDDQREQLSVAEEMDSYSATVEEVATA 285
QY 175 TELARYASDQALDLSGLT-----SFDLIQTALLOSVAANNKAAELKEMQNPVVP 228
DB 286 QSVADTAADTTVA-TAGKQTAEDAIDAIDAVQETMQTTVANVDALEDLTTEIDD----- 339
QY 229 GKTPAIAQSLVDQTDAT-----ATQIEKQGNATGD--AYFAGQNASGAVENAKSN 276
DB 340 -----IAELISDIARQTNLALNANIEARAGSGGTGNGDFAVVAVDEVKELATESQRS 394
QY 277 NSISN-IDSAAKAIATATQTAIAQKFPDPSI-----LQAE 313
DB 395 KDIAELIEVQSQTATTVEIRVABQVRVNDGAAVEETVDAFQAVTENIQETTDGQVEIS 454
QY 314 Q-----MVIQAEKDLNKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLLDDA 363
DB 455 QAMDEQAQSRVSVSSVDDIATISQATADRAEN---VSNASEEQATSEITVETGSLQSLAA 511
QY 364 ENETASILMSGFR-QMIMFNTENPDQAAQOE 395
DB 512 QADTLEDRINEFRTEATGHTGERTDAPAGQSD 544

RESULT 15

T34852

probable secreted protein - Streptomyces coelicolor

C:Species: Streptomyces coelicolor

C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999

C:Accession: T34852

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, February 1999

A:Reference number: Z12559

A:Accession: T34852

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1156 <OLI>

A:Cross-references: EMBL:AL035478; PIDN:CAB36606.1; GSPDB:GN00070; SCOEDB:SC2G5.19

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC2G5.19

Query Match 7.0%; Score 158.5; DB 2; Length 1156;
Best Local Similarity 21.2%; Pred. No. 0.54; Indels 65; Gaps 15;
Matches 99; Conservative 75; Mismatches 227;
QY 12 ETERPPADLSAAGLEASAAKSAEARIAGAEAKPKESKTDSDVERWILRSVAVNALMS- 70
DB 367 DASKTKAARLAAGAGARNAAKARKAAQAATAQAAAAAGISAAATARDSAAAAQA 426
QY 71 --LADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTAQATQATYDTIFTSTSLA 128
DB 427 AVAQASGAAGAAQSEAAVAAAAAADAQAARA---TKAANRAQSLAATAASAAAAARKAA 482
QY 129 DTQAA-LVSLQDAVNTIKDTAATDETAIAAE-WETKNADAIVKGAQITELAKYASDNOA 186
DB 483 DSAHAHAKEAADAADAADAAGEADDYANKAKAWA---ADSVAA-----AELAAKAVDDAR 535
QY 187 ILDSLGKITSFDLLQTLQSVANNKAAELKEMQNPVVPKTPAIAQSLVDQTDATA 246
DB 536 AVEAAAAEAAEKLAHDTEQSLAEARENA--AAEAEDREAAARN-----AATEADRLDAQT 588
QY 247 TQIEKQGNATGDAYFAGQNASGAVENAKSNNSISNIDSAKAIATATAKQIAEAQKFPDS 306
DB 589 ----KDFISRAEAAYASGDTASALANGR-----KAANLISITIGTWSRAAAEY 633
QY 307 PIQPAEQMVIQAEKDLNKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLLDDAENE 366
DB 634 ALAGEDEVLTWVSTDRQIAQRQDDSE---TALAVANVSTEAVANAAAAAIAENSDDPQS 688
QY 367 TASILMSGFRQ-----HMFNTENPDQAAQOEAAQARAARAAAGDDSDAAAAALADA 418
DB 689 VRNFLTGTGHEAAALDRVRDILRIILGDNPK-----AVKDAQAALDDGSPSALHAF 740
QY 419 OKAL--EAALGKAGQOQGIILNALGQIASAAVVSAGVLPLOQVLR 462
DB 741 FRALPDKAALDDRAVILITILNTAGPYTAAQAQVA-----LEGTSWNR 782

Search completed: February 7, 2002, 21:38:09

Job time: 7098 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:42 ; Search time 76.51 seconds
(without alignments)
221.877 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MYNPIGPGIDERTPPAD.....SAAVVSAGVLPQLQVLWIRA 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 segs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	188	8.3	1120	1 STFR_ECOLI	P76072 escherichia
2	174.5	7.7	545	1 HTR5_HALN1	Q48318 halobacteri
3	158	7.0	535	1 HTR1_HALN1	P33741 halobacteri
4	158	7.0	810	1 HTR4_HALN1	Q9hp84 halobacteri
5	154.5	6.8	763	1 HTR4_HALSA	Q48317 halobacteri
6	154.5	6.8	764	1 HTR2_HALN1	Q9hp81 halobacteri
7	154.5	6.8	764	1 HTR2_HALSA	P71410 halobacteri
8	152	6.7	1561	1 SPAP_STRMU	P23504 streptococc
9	151.5	6.7	1561	1 HRPX_PSESY	P41501 pseudomonas
10	150.5	6.7	1528	1 SPAA_STRDO	P21979 streptococc
11	148	6.5	505	1 FLJB_SALTY	P52616 salmonella
12	148	6.5	881	1 YJH8_YEAST	P21979 streptococc
13	147	6.5	778	1 HTR6_HALSA	Q48319 halobacteri
14	146.5	6.5	774	1 STF_LAMBD	P03764 bacterioph
15	146	6.5	1714	1 SYEP_DROME	P28668 drosophila
16	145	6.4	705	1 CWBA_BACSU	Q02113 bacillus su
17	145	6.4	1565	1 PAC_STRMU	P11657 streptococc
18	144	6.4	1637	1 MRSP_STAAU	P80544 staphylococ
19	143.5	6.3	1306	1 MSB2_YEAST	P32334 saccharomyc
20	142.5	6.3	778	1 HTR6_HALN1	Q9hr92 halobacteri
21	139	6.1	797	1 VGLX_HSVB	P28968 equine herp
22	138	6.1	1609	1 LMGI_HUMAN	P11047 homo sapien
23	138	6.1	2843	1 APC_MOUSE	P25054 homo sapien
24	137.5	6.1	2453	1 NCRI_MOUSE	Q60974 mus musculu
25	137.5	6.1	2492	1 TALA_DICDI	P54633 dictyosteli
26	137	6.1	500	1 FLJB_SALAE	P52615 salmonella
27	137	6.1	535	1 HTR1_HALSA	P33955 halobacteri
28	137	6.1	1607	1 LMGI_MOUSE	P02468 mus musculu
29	136	6.0	2541	1 TALI_MOUSE	Q60399 mus musculu
30	136	6.0	2845	1 APC_MOUSE	Q61315 mus musculu
31	135.5	6.0	1061	1 TRC4_ECOLI	P27189 escherichia
32	135.5	6.0	1969	1 MYSA_CAEEL	P12844 caenorhabdi
33	135	6.0	1379	1 YFF9_SCHPO	O14066 schizosacch

RESULT 1

STFR_ECOLI
ID STFR_ECOLI STANDARD; PRT; 1120 AA.
AC P76072: P77560;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC.
GN STFR OR B1372.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick J.A., Goeden M.A., Rose D.J., Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251357; PubMed=9097039;
RA Aliba H., Kashimoto K., Kimura S., Kitakawa M., Kitagawa M., Kasai K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K., Nakano K., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Nakade S., Seki Y., Sivasubraman S., Tagami H., Takeda J., Sempel G., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT "A 570-kb DNA sequence of the Escherichia coli K-12 genome corresponding to the 28.0-40.1 min region on the linkage map";
RNA Res. 3:363-377(1996).
CC [1] SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AE000234; AAC74454.1; ALT_INIT.
DR EMBL; D90774; BAA14966.1; -;
DR EMBL; D90775; BAA14975.1; -;
DR HSP; P04002; IWFA.
DR Ecogene; EG13370; stfr.
DR InterPro; IPR000122; Chemotaxis_transducer.
KW Hypothetical protein; Fiber protein; Repeat; Complete proteome.
SQ SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

ALIGNMENTS

34 134 5.9 564 1 M12_STRPY
35 134 5.9 1036 1 NIT2_NEUCR
36 134 5.9 2842 1 APC_FAT
37 132.5 5.9 573 1 YEY8_YEAST
38 131.5 5.8 1065 1 SED4_YEAST
39 131.5 5.8 1411 1 TCOF_HUMAN
40 130.5 5.8 3726 1 TRX_DROME
41 130 5.7 475 1 MOTD_RHIME
42 130 5.7 490 1 CE05_ECOLI
43 130 5.7 1230 1 ST20_CANAL
44 129 5.7 1500 1 SSP5_STRGN
45 128.5 5.7 534 1 HTR2_NATPH

P19401 streptococc
P19212 neurospora
P70478 rattus norv
P40095 saccharomyc
P25365 saccharomyc
Q13428 homo sapien
P20659 drosophila
Q52964 rhizobium m
O47500 escherichia
Q92212 candida alb
P16952 streptococc
P42259 natronomona

Query Match 8.3%; Score 188; DB 1; Length 1120;
Best Local Similarity 23.0%; Pred. No. 0.012;
Matches 103; Conservative 68; Mismatches 203; Indels 74; Gaps 13;

QY 8 GPIDETERTPPA-----DLSAQGLEASAAKSAEQAORTAGAEAKPKESKTSVERWSILRS 63
DB 91 GAMEDDARPEALRRFELM---VEEVARNASAVAQNTAAAKSKASDASTAREAAATHAAD 147
QY 64 AVNALMSLADKLGIASSNSSTSRSDVSTTATATPTPPPTSDDYKTOAQAYDIFT 123
DB 148 AADSARAASTSAGQAAS-SQAASSSAGTASTKATEASKSAAAEESKSAATSAAGAKT 206
QY 124 STSLADIQAALVSLQDAVNTIKDPAETAEATAAEWETKNADAIVKGAQITELAKYASD 183
DB 207 SETNAS-----ASLQSAATSAATTKASEAATSA-----RDAASAKAEAKSSSETNASSSA 257
QY 184 NQALDLSGLKLTSPDLQTTALLOSVANNKAAELLKEMQNDPVVPGKTPAIAQSLVDQTD 243
DB 258 SSAASSATAAGNSAKAAKTS-----ETNARSSETAAGQSSASAAAGSKTAAASSASAAS 312
QY 244 ATATQIEKDGNAIGDAYFAGONAGAVENAKSNNSINISDAKAAIATAKTQIAEAAOKKF 303
DB 313 A-----GQASASATAAGKSA-----ESAASSASTATTKAGEATEQ- 347
QY 304 PDSPILOEAEQWVIOAEKDLKNIPKADGSDVPNPCTTTVGSKQOGSSIGSIRVSMLLDDA 363
DB 348 -----ASAAARSAAKTSSETNAKASSETSAESSKTAASASSASSASSASASAK--DEA 400
QY 364 ENETASILMSGFOMIHFNTENPDSQAQOELAAQA-----RAAKAAGDSSAAALAD 417
DB 401 TRQASAKSSATTAATKATEAAGSATAAAGSKSTAESAATRAETAARAEADIASAVALD 460
QY 418 AOKALEAALRAGQOOGILNALGQIASA 445
DB 461 ASTT-----KKGIV-----QLSSA 474

RESULT 2

HTR5_HALN1 STANDARD; PRT; 545 AA.
ID HTR5_HALN1
AC Q48318; Q9HP85;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE HALOBACTERIAL TRANSDUCER PROTEIN V.
GN HTR7 OR HTPV OR VNGI1759G.
OS Halobacterium sp. (strain NRC-1), and
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091, 2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithausen B., Kellier K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlischer M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
RT "Genome sequence of Halobacterium species NRC-1".
RT Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC SPECIES=H.salinarum; STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RA Oesterheilt D.
RT "A family of halobacterial transducer proteins".
RT FEMS Microbiol. Lett. 139:161-168(1996).

CC -1- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
TRANSDUCTION.
CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).

CC EMBL; AE005080; AAG19985.1; -
DR EMBL; X95589; CAA64842.1; -
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00015; MCPs; signal; 1.
DR SMART; SM00304; HAMP; 1.
DR SMART; SM00283; MA; 1.
KW Transducer; Transmembrane; Complete proteome.
FT TRANSMEM 10 30 POTENTIAL.
FT TRANSMEM 44 64 POTENTIAL.
FT TRANSMEM 80 100 POTENTIAL.
SQ SEQUENCE 545 AA; 57070 MW; FDD870389C2F428B CRC64;

Query Match 7.7%; Score 174.5; DB 1; Length 545;
Best Local Similarity 21.7%; Pred. No. 0.024;
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;

QY 12 ETEPTPADLSAQGLEASAAKSAEQAORTAGAEAKPKESKTSVERWSILRSVNALMSL 71
DB 118 EEERA-----EAERAREKAEQKQAEER--QTAESAESAKODARESAEIBOLAADLESQ 169
QY 72 ADKLGIASSNSSTSRSDVSTTATATPTPPPT-SDDYKTOAQAYDIT-----FTST 125
DB 170 ATEVG-ATLEASGDDITARVDAITDNEAEVATVYVNDMLTTWERTIDEIQGFSTVTT 228
QY 126 SLADIQAALVSLQDAVNTIKD-----AATDEE-----TAIAAEWETKNADAIVKGAQITE 176
DB 229 ASREATAGAEIQDASQTVSESVQIEAAGTDDQEQLESVAEEMDSYATVEEVAATQAS 288
QY 177 LAKYASDNOAILDSLGLT-----SFDLQTTALLOSVANNKAAELLKEMQNDPVVPGK 230
DB 289 VADTAADTTDVA-TAGKQTAEDAIDAIVQETMOTTVANNVDALEDLTTEIDD----- 340
QY 231 TPAIAQSLVQDQDAT-----ATQIEKDGNAIGD--AYFAGONASGAVENAKSNNS 278
DB 341 ---IAELISDIAEQTNMALNANTEAARAGSGGSGNGDGFVAVADEVKELATESQSAKD 397
QY 279 ISN-IDSAAKAAITAKTQIAEAKKFPDPSPILOEAEQWVIOAEKDLKNIPKADGSDVPNP 337
DB 398 IAEELIEVQSQATTVEEI-----RYAEQRVNDGAAAEVETVDAFGAVTENI 444
QY 338 GTTVGSKQOGSSIGSIRVSMLLDADNETASILMSGFOMIHFNTENPDSQAQOELAA 397
DB 445 QETTDGVQE-----ISQAMDEQAQRSESVSS-----VDDIATISQATA 483
QY 398 AQARAARAAAGDSSAAALADAQAALKAALGKAGQOOGILN 437
DB 484 DRAENVSAASEEQ-TASITEVTSLSQSLAAQAQADLEDRLN 522

RESULT 3

HTR1_HALN1 STANDARD; PRT; 535 AA.
ID HTR1_HALN1
AC P33741; Q9HPF6;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE SENSORY RHODOPSIN I TRANSDUCER (HTR-1) (METHYL-ACCEPTING PHOTOTAXIS
DE PROTEIN I) (MPP-1).
GN HTR1 OR HTRI OR VNGI1659G.

OS Halobacterium sp. (strain NRC-1), and
 OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091, 2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=NRC-1;
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sroogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weiti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
 RA "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10; 350-377 AND 457-476.
 RC SPECIES=H.halobium; STRAIN=FLX5R;
 RX MEDLINE=93101637; PubMed=1465418;
 RA Yao V.J., Spudich J.L.;
 RT "Primary structure of an archaeobacterial transducer, a
 RT methyl-accepting protein associated with sensory rhodopsin I";
 RL Proc. Natl. Acad. Sci. U.S.A. 89:11915-11919(1992).
 CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
 CC SENSORY RHODOPSIN I (SR-I) TO THE FLAGELLAR MOTOR. RESPONSES TO
 CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
 CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE005075; AAG19913.1; -
 DR EMBL; L05603; AAA72315.1; -
 DR PIR; A47190; A47190.
 DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR000658; DUF5.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00672; DUF5; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR SMART; SM00304; HAMP; 2.
 DR SMART; SM00283; MA; 1.
 KW Transducer; Photoreceptor; Transmembrane; Methylation.
 FT INIT_MET 0 0
 FT DOMAIN 1 13 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 14 28 POTENTIAL.
 FT DOMAIN 29 38 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 39 54 POTENTIAL.
 FT DOMAIN 55 535 CYTOPLASMIC (POTENTIAL).
 FT MOD_RES 265 265 METHYLATION.
 FT MOD_RES 272 272 METHYLATION.
 FT MOD_RES 279 279 METHYLATION.
 FT MOD_RES 463 463 METHYLATION.
 FT MOD_RES 472 472 METHYLATION.
 SQ SEQUENCE 535 AA; 56544 MW; B9945E4F66A9D091 CRC64;
 Query Match 7.0%; Score 158; DB 1; Length 535;
 Best Local Similarity 18.7%; Pred. No. 0.17;
 Matches 88; Conservative 83; Mismatches 189; Indels 110; Gaps 15;
 QY 8 GPIDETERTPADLSAQGLESAANKSAE--AQRTAGAEAKPK-----SKTDSVERW---- 58
 DB 103 GRLEMET-RADLEETQAEAREAEQAQAEAREAREALATYQDTAKRYGETM 161

QY 59 -----SILRSVNALMSLADKLGIASSSSST 86
 DB 162 EAAATGDLTQRVDVTDHEAMETVCTAFNQMDLLOATVTVTTVADIEIAKTERMSSET- 220
 QY 87 SRSADVST---TATAPTPPPPTSDYKTOAQATYDTTFTSTSLADIAQALVSLQDVTN 143
 DB 221 --SADIEASAGDTVEAVSKIESQANDORTELSAAD-----DVQVSASAEIAAT 269
 QY 144 IKDPAATDEETAIAE-----WETNDAIKVGAOITELAKYASDNOAIL 188
 DB 270 IDDLASREDVATSDAARDSSKALDEMSSIEETEVDVAGVQVEQLDQVREITDIVDI 329
 QY 189 DSLGLKLTDFLLQTLQSVANN-----KAEELKEMQDNVFWPKTPAIQAS 237
 DB 330 TEIGQTNMALNASIEAARAGGNADGGFSVVADEVKDLAEETQDR---ANEIAAVVEK 386
 QY 238 LVQDTATATQIEKDGNAGIDAYFAGQNASGAVENAKNSISNTSDSKAIAIATAKTOIA 297
 DB 387 VTAQTDVYASIQQTRTRV-----ESGSETVE-----STLRDITRTADSIAEVNSID 434
 QY 298 EAQKFPDPSILQEAQMVIOAEKDKNKKPADGSDVPNGTGVGSKQGGSSIGSIRVS 357
 DB 435 EIQRITTSQAETVQSTATSVERVAGLSDDTTALASDAE---SAVIGQRESAEETIAA- 487
 QY 358 MLLDDAENETASILMSGFRQIMHMENTENPDQSQAQQLAAQARAAGAAG 407
 DB 488 -SLEQFQNTAVEQLQS-----RVASFTVATEDSETAGGSVEQPMVRGAGD 532
 RESULT 4
 HTR4_HALN1 STANDARD: PRT; 810 AA.
 ID HTR4_HALN1
 AC Q9HP84;
 DT 20-AUG-2001 (Rel. 40, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HALOBACTERIAL TRANSDUCER PROTEIN IV.
 GN HTR5 OR HTPIV OR VNGI760G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sroogna J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Weiti R., Goo Y.A.,
 RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.,
 RA "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
 CC TRANSDUCTION (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AE005080; AAG19986.1; -
 DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR000658; DUF5.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00672; DUF5; 1.
 DR Pfam; PF00015; MCPsignal; 1.

```

RESULT 5
HTR4_HALSA
ID HTR4_HALSA STANDARD; PRT; 810 AA.
AC Q48317:
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT DT DT
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE HALOBACTERIAL TRANSDUCER PROTEIN IV.
GN HTR5 OR HTPIV.
OS Halobacterium salinarium.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=S9;
RX MEDLINE=96275896; PubMed=8674984;
RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
RT Oesterheld D.;
RA "A family of halobacterial transducer proteins."
RL FEBS Microbiol Lett. 139:161-168(1996).
CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION.
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC

```

OS Halobacterium sp. (strain NRC-1).

CC -|- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORIAL TRANSDUCERS FAMILY.
CC -|-
CC

Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;
 Halobacterium.
 NCBI_TaxID=64091;
 [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Shrogha J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR SENSORY
 CC RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO LIGHT
 CC CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION (BY
 CC SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; AEO05080; AAC19989.1;
 DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR000658; DUF5.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00672; DUF5; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR SMART; SM00304; HAMP; 2.
 DR SMART; SM00283; MA; 1.
 KW Transducer; Photoreceptor; Transmembrane; Methylation;
 KW Complete proteome.
 FT INIT_MET 0 BY SIMILARITY
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 16 36 POTENTIAL.
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 278 297 POTENTIAL.
 FT DOMAIN 298 763 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 763 AA; 78911 MW; CF7A9FF04DFE309A CRC64;
 Query Match 6.8%; Score 154.5; DB 1; Length 763;
 Best Local Similarity 22.6%; Pred. No. 0.39;
 Matches 105; Conservative 71; Mismatches 196; Indels 93; Gaps 18;
 Oy 8 GPIDETERTPADLSAQGLEASANKSAE-----AQRIAGAEAKPKE---SKTDSVERW 58
 Db 336 GTLAESFRSMDLSSESUTDERATAREDAEDAEQQRADAEAREDAEAKDAQTA 395
 Oy 59 SILRSV-----NALMSLADKLGIASSNSSSTSRSA-----DVIDSTTATAPT 101
 Db 396 RALESAAADYEALTAVAD--GDLTRRVYDASRDHDAMARIGHALNDMLDIETSVAAATA 453
 Oy 102 PPPPTSDYKTAQATYDTITFTSLADIQAALVSLQDAVNTKIDTADEETAIAEWE 161
 Db 454 FSDHVSAAQAEVADAGDAIDAGT---DVSFVAVDEISDGAET---QTDRLHEVAGEVD 505
 Oy 162 TKNDATAIKVGAQITELAKYASDNOAILDLSGLKLTSEDLLOTALLOSVAANNKAAELKEM 221
 Db 506 DLSASAEVATVSLADTAQAASAVDD--GRQATEDRVER--MDDVADAEARAADAMA 562
 Oy 222 QDNVPVP--GKTPAIAQSLVDQTDATQIEKDGNAIGDAYFAGONASGAVENA----- 273
 Db 563 LDSEMADIGEIVDVADIADQTNMLAL-----NASIEAARTGADGDFVAVADEVKTLA 616

274 -KSNNSISNIDSAAKATATATQTAIAEQKFPDPSPILQEAQMVIQAEKDLKNKPADGS 332
 Db 617 EESRDAAEEDIESRLALQGVSDVADEMRASTSDT--VSDGRATVGDATALLDDV----- 668
 Oy 333 DVPNPGTTVGSKQGGSSIGSIRVSMMLDDDAENETASTILMSGFQMIHMENTENPDSQAA 392
 Db 669 -----VSFVADTDTAAGEIRAA---TDQAHAAASRVASAVDEV-----AGI 706
 Oy 393 QOELAAQARA-AKAAGD-----DSAAALADAQAQKALEAALGK 428
 Db 707 SQETAAQAQAVADSAAATQTDTLSSVDDAAADLADRAAALDDLLAE 751
 RESULT 7
 HTR2_HALSA STANDARD; PRT; 764 AA.
 ID HTR2_HALSA
 AC P71410;
 DT 15-JUL-1998 (Rel. 36, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SENSORY RHODOPSIN II TRANSDUCER (HTR-II) (METHYL-ACCEPTING PHOTOTAXIS
 DE PROTEIN II) (MPP-II).
 GN HTR2 OR HTRII.
 OS Halobacterium salinarum.
 OC Archaea: Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OC NCBI_TaxID=2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=FLX15;
 RX MEDLINE=96323203; PubMed=8710852;
 RA Zhang W., Brooun A., Mueller M.M., Alam M.;
 RT "The primary structures of the Archaeon Halobacterium salinarum blue
 RT light receptor sensory rhodopsin II and its transducer, a methyl-
 RL accepting protein";
 RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).
 CC -1- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
 CC SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONDS TO
 CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
 CC -1- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U62676; AAC44369.1;
 DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR000658; DUF5.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00672; DUF5; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR SMART; SM00304; HAMP; 2.
 DR SMART; SM00283; MA; 1.
 KW Transducer; Photoreceptor; Transmembrane; Methylation.
 FT INIT_MET 0 BY SIMILARITY
 FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 16 36 POTENTIAL.
 FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 278 298 POTENTIAL.
 FT DOMAIN 299 764 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 764 AA; 79187 MW; 1E0D7B4E460FC588 CRC64;

Query Match 6.8%; Score 154.5; DB 1; Length 764;
 Best Local Similarity 22.6%; Pred. No. 0.39;
 Matches 105; Conservative 71; Mismatches 196; Indels 93; Gaps 18;
 Oy 8 GPIDETERTPADLSAQGLEASANKSAE-----AQRIAGAEAKPKE---SKTDSVERW 58
 Db 563 LDSEMADIGEIVDVADIADQTNMLAL-----NASIEAARTGADGDFVAVADEVKTLA 616

Db 337 GTLAESFRMRDLSLESLDAERATARAEDAREDAEQORADAEAAAREDAEFAARKDAQETA 396
QY 59 SILRSVAV-----NALMSLADKLGTASSNSSSTSRSA-----DVSSTTAPT 101
Db 397 RALESAAADYEALUTAVAD--GDLTRVYDASRDHDMARIGHALNDMLDDIETSVAAATA 454
QY 102 PPPPTSDDYKTAQTAYTFTSTSLADIAQALVSLQDVAVTNKDSTADEETAEIAAEWE 161
Db 455 FSDHVSDAARVEADAGDAIDAGT---DVSSTAVDEISDGAETE---QDRLHEVAGEVD 506
QY 162 TKNADATKVAQAQITELAKYASNDQAILDSGLKTSFLLQTLALLOSVAANNKAAELKEM 221
Db 507 DLSASAEVETVASLADTAQAAASAVDD--GROATEDAVET--MDDVADDAEAAADAMD 563
QY 222 QDNVPVP--GKTPATAQSIQVDTATATQIEKDGNAIGDAYFAGNASSGAVENA----- 273
Db 564 LDSEMADEIGELVDAIADOTNMLAL-----NASTEAARTGADGGFAVVADEVKTLA 617
QY 274 -KSNNSISNDSAKAATATAKTQIAEAQKPPDSPILQEAQEMQVIAQEKDLKNIKPADGS 332
Db 618 EESRDAEDIESRLALQGVSDVADEMRASTDT--VSDGRATVGDAAATLDDV----- 669
QY 333 DVPNPGTVGSKQKQSGSIGSVIRVMSLDDDAENETASILMSGFQMTIHMENTNPDSQAA 392
Db 670 -----VSFVADTDTAAGQIRAA---TDQHAASRVASAVDEV-----AGI 707
QY 393 QOELAAQARA-AKAAGD-----DSAAALADAQKALEAALGK 428
Db 708 SOETAQAQATVADSAAATQDTLSSVDDAAADLADRAALDDLLAE 752
RESULT 8
SPAP_STRMU
ID SPAP_STRMU STANDARD; PRT; 1561 AA.
AC P23504;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE ANTIGEN I/II PRECURSOR.
GN SPAP.
OS Streptococcus mutans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1309;
[1]
RN SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN-NG5 SEROTYPE C;
RX MEDLINE=90076473; PubMed=2687020;
RA Kelly C., Evans P., Bergmeier L., Lee S.F., Progulsk-Fox A.,
RA Harris A.C., Aitken A., Bleiweis A.S., Lehner T.;
RT "Sequence analysis of the cloned streptococcal cell surface antigen
RT I/II".
RL FEBS Lett. 258:127-132(1989).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-NG5 SEROTYPE C;
RX MEDLINE=91207143; PubMed=1982405;
RA Kelly C., Evans P., Ma J.K.C., Bergmeier L.A., Taylor W., Brady L.J.,
RA Lee S.F., Bleiweis A.S., Lehner T.;
RT "Sequencing and characterization of the 185 kDa cell surface antigen
RT of Streptococcus mutans".
RL Arch. Oral Biol. 35:33S-38S(1990).
CC -!- FUNCTION: SURFACE PROTEIN ANTIGEN IMPLICATED IN DENTAL CARIES.
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN; CELL WALL.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X17390; CAA35253.1; -
DR PIR: S06839; S06839; Gram_pos_anchor.
DR InterPro: IPR001899; Gram_pos_anchor.
DR Pfam: PF00746; Gram_pos_anchor; 1.
DR PROSITE: PS00343; GRAM_POS_ANCHORING; 1.
KW Antigen; Signal; Transmembrane; Repeat; Cell wall; Dental caries.
FT SIGNAL 1 38
FT CHAIN 39 ? CELL SURFACE ANTIGEN I.
FT CHAIN 997 1561 CELL SURFACE ANTIGEN II (PROBABLE).
FT DOMAIN 39 1536 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 1537 1556 MEMBRANE ANCHOR.
FT DOMAIN 1557 1561 CYTOPLASMIC (POTENTIAL).
FT DOMAIN 60 550 HELICAL (POTENTIAL).
FT DOMAIN 219 464 3 X TANDEM REPEATS, ALA-RICH.
FT DOMAIN 847 963 3 X TANDEM REPEATS, PRO-RICH.
FT DOMAIN 1528 1533 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT PROTEINS.
SQ SEQUENCE 1561 AA; 170060 MW; 540D92768FC8AB4B CRC64;
Query Match 6.7%; Score 152; DB 1; Length 1561;
Best Local Similarity 22.3%; Pred. No. 1.2;
Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;
QY 48 KESKTDVSVERWSILRSVAVNALMSLADKLGTASSN-SSSSTSRSDVDSTTATATPTPPPT 106
Db 2 KVRKTYGFRKSKITLCGAVLGTVAASVAGQKVFADETTTSDVDTKVVGQTGTGNPAT 61
QY 107 -----SDDYKTAQATAYDTFTST-----SLADIAQAALVS-LQDAVTNPKDT 147
Db 62 NLPEAQGSASKQAEQSQTCLEROMVHTIEVPKTDLDQAADKASGAVVVQDADVN-KGT 120
QY 148 AATDEETAAAEWE-----TKNADAIAK-----VGAQITELAKYASNDQAILDSLGK 193
Db 121 VKTAAE-AVOKETEIKEDYTKQAEIDIKKTTDOYKSDVAAHEAEVAKIKAKNQATKEQYK 179
QY 194 LTSFDLL-QTALLGQSVANNKAAEL-----LKEMODNPVVPKTPALQAQSLVDQTDATAT 247
Db 180 ----DMVAHKAEEVERINAANAASKTAYEAKLAQYQADLAQVKTNAANQASY-QKALAA 234
QY 248 QIE-----KDGNAIGDAYFAGNASSGAVENAKSNNSISNIDSAKAAIATATK-----TQIAEA 299
Db 235 QAEILRVQEAANAATAAAY---DTAVAANNAKNTETIAAANEIEIRKRNATAKAEYETKLAQY 291
QY 300 Q---KKFPDPSPIQEAQ-QMVIQA-EKDLKNIKPAD-----GSDVPNPGTVG 342
Db 292 QAEILRVQEAANAATAAAYQAKLTAYQTELARVQKANADAKAAAYEAAVAAANNKNAALUTAE 351
QY 343 GS-----KQOQSSIGSVIRVMSLDDDAENETASILMSGFQMTIHMENTE 385
Db 352 NTAIKQRENAKATYEAALQYQYEAADLAQVKKANNAANADYQAK---LTAYQTELARVQKA 408
QY 386 NPDSQAQAEQELAAQARAATAA---GDDSAQAAL-ADAQKALEAALGK 428
Db 409 NADAKAAEAAVAAANNAANAALTAENTAIAKRNADAKADYEAALAK 454
RESULT 9
HRPK_PSESY
ID HRPK_PSESY STANDARD; PRT; 641 AA.
AC P41501;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-NOV-1995 (Rel. 32, Last annotation update)
DE PATHOGENICITY LOCUS PROTEIN HRPK.
GN HRPK.
OS Pseudomonas syringae (pv. syringae).
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=321;

```

RN RP SEQUENCE FROM N.A.
RC STRAIN-PSS61;
RX MEDLINE=94148760; PubMed=8106313;
RA Xiao Y., Heu S., Yi J., Lu Y., Hutcheson S.W.;
RT "Identification of a putative alternate sigma factor and
RT characterization of a multicomponent regulatory cascade controlling
RT the expression of Pseudomonas syringae pv. syringae Pss61 hrp and
RT hrmA genes.";
RL J. Bacteriol. 176:1025-1036(1994).
RN RP SEQUENCE FROM N.A.
RC STRAIN-PSS61;
RX MEDLINE=94100578; PubMed=8274770;
RA Heu S., Hutcheson S.W.;
RT "Nucleotide sequence and properties of the hrmA locus associated with
RT the Pseudomonas syringae pv. syringae 61 hrp gene cluster.";
RL Mol. Plant Microbe Interact. 6:553-564(1993).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U03855; AAA17653.1; --
SQ SEQUENCE 641 AA; 67678 MW; EC0989941E5B46C8E CRC64;

Query Match 6.7%; Score 151.5; DB 1; Length 641;
Best Local Similarity 21.5%; Pred. No. 0.44;
Matches 90; Conservative 56; Mismatches 181; Indels 91; Gaps 15;

QY 53 DSVRWSTLSRAVNAIADKLGIASSNSSSTSRADVDS-----TTATPTPPPT 106
DB 53 DSSPSFGTEQSGSLMSL-----LTRSSSESTSS-SVQDSQDVSPMTSVSTASAPT 106
QY 107 SDDYKTAQATYDPTFTSTSLADIQAALVSLQDAVTNI---KDTAATDEETALAAEWET 162
DB 107 AASNAPANASATDAFLDNSEYSPALKRWEPVAVNLPPEERQAAKELNRPAAAWMA 166
QY 163 KNADAIVKQAQITELAYASDNQAILDSL-----GKLTSDLLQALLOSVAANN 212
DB 167 R-----ENGPNAERAMAFINANPALKTAVDVGKDGNGADGKITNKLKAFKNMEKAADN 221
QY 213 KAELLKEMODN-----PVVPEKTPAIAQSLVDQDTATATQIEK 251
DB 222 AKDVAKYMEDNPGADPQSLQEMVRSAAVMRANMPLATAADPHHVAAGADKTDV----- 274
QY 252 DQNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKQIAEAKKFPDPSILQE 311
DB 275 DGNV-----SAEGLKALIKSNPGLSG-----TLKQSSNNWSQAGFLSQ 312
QY 312 AEQWVIAQEKDLNLIKPADGSDVPNPQTTVGGSKQGGSSGSIKRVSMLLDQAE-NETASI 370
DB 313 VDEAGLTGRKKAH-SPDQVDFASNMSEWI---RKSAPKNGGQFPASMLSDAATLNSVAGI 368
QY 371 LMSGFQMIHMFENTENPDSQAQOELAAQARAQAKAGDDSAARAALADAQKALEALCK 428
DB 369 DISKLNQVFE---EKPKATYGAQAAVMIKLQQTQOSVIAGRDLRNRTEK-EGALNE 421

RESULT 10
SPAA_STRDO
ID SPAA_STRDO STANDARD; PRT; 1528 AA.
AC P21979;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE ANTIGEN I/II PRECURSOR.
GN SPAA.

Streptococcus downei (Streptococcus sobrinus).
Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
NCBI_TaxID=1317;
[1]
SEQUENCE FROM N.A.
STRAIN=6715 / SEROTYPE G;
MEDLINE=91310320; PubMed=1855987;
Lapolla R.J., Haron J.A., Kelly C.G., Taylor W.R., Bohart C.,
Hendricks M., Pyati J., Graff R.T., Ma J.K.C., Lehner T.;
"Sequence and structural analysis of surface protein antigen I/II
(Spaa) of Streptococcus sobrinus.";
RL Infect. Immun. 59:2677-2685(1991).
RN RP SEQUENCE OF 423-817 FROM N.A.
RX MEDLINE=90299827; PubMed=1694526;
RA Goldschmidt R.M., Thoren-Gordon M., Curtiss R. III;
RT "Regions of the Streptococcus sobrinus spaa gene encoding major
RT determinants of antigen I.";
RL J. Bacteriol. 172:3988-4001(1990).
CC -!- SUBCELLULAR LOCATION: TYPE I MEMBRANE PROTEIN. CELL WALL.
CC -!- MISCELLANEOUS: IMMUNODOMINANT DETERMINANTS ARE LOCATED IN THE
CC C-TERMINAL TWO-THIRDS OF THE SPAA PROTEIN.
CC -!- SIMILARITY: BELONGS TO THE SPAP/SSP5/SPAA FAMILY.
CC -!- SIMILARITY: TO OTHER STREPTOCOCCAL AND STAPHYLOCOCCAL PROTEINS
CC IN THE REGION OF THE MEMBRANE ANCHOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X57841; CAA00973.1; --
EMBL; M38210; AAA26977.1; --
DR InterPro; IPR001899; Gram_pos_anchor.
DR PROSITE; PS00343; GRAM_POS_ANCHORING. 1.
KW Antigen; Signal; Transmembrane; Repeat; Cell wall.
FT SIGNAL 1 ? ? POTENTIAL..
FT CHAIN ? ? CELL SURFACE ANTIGEN I.
FT CHAIN ? 1528 CELL SURFACE ANTIGEN II.
FT DOMAIN 1503 1508 CONSERVED IN GRAM-POSITIVE COCCI SURFACE
FT FT PROTEINS.
FT FT CONFLICT 427 427 A -> E (IN REF. 2).
FT FT CONFLICT 431 431 Q -> K (IN REF. 2).
FT FT CONFLICT 434 434 Q -> K (IN REF. 2).
FT FT CONFLICT 531 531 A -> S (IN REF. 2).
FT FT CONFLICT 600 600 A -> S (IN REF. 2).
SQ SEQUENCE 1528 AA; 165572 MW; B235F9CCD92838E0 CRC64;

Query Match 6.7%; Score 150.5; DB 1; Length 1528;
Best Local Similarity 23.5%; Pred. No. 1.4;
Matches 110; Conservative 68; Mismatches 192; Indels 99; Gaps 22;

QY 38 QRIAGAEAKPKESKTSQVRSILRSVAVNAIADKLGIASSNSSSTSS-RSADVST 95
DB 16 KRUGA-AKVSGRTLS---GALLGTAI--LASAGQALAEETSTSTSGDFAVVGTE 69
QY 96 TATAPTPTPTSDDYKTAQATAY-----DTIFTSTSLADIAQALVSLQDAVTNFKDTA 148
DB 70 TGNPATNLPDKQDNPPSSQAEISQAQARQKTCGMSVDVSTSELDEAKSPQEGVTVSODA 129
QY 149 ATDEETAIAAEWETKNADAIK--VGAQITELAKYASD-----NQAILSLGKLTGFDL 199
DB 130 TVNKGTVPEPSDEANQKEPEIKDDYSKQAAIDQKATEDYKASVANAQAEORINQIEIAKK 189
QY 200 LQYALLOSVAANNKAAE--LLKEMODNPVVPKTPAIAQSLVDQDTATATQIEKDGNAIG 257
DB 190 AQYE--QDLAANKAEVERSLMRKRPRIYEAK---LAQNKQDL--AAIQANSDSQA-- 240

```

```
QY 258 DAYFAGON-----ASGAVENAKSNNSISNIDSAKAAATATKTAQTAENAKK 302
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 241 -AYAAKEAYDKWARVQAANAANKAYEEAALAAANTAKN-DQIKAEIEAIQORSAGA--- 295
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 303 FPDSPILQEAQOMVIOAEKDLKNIKPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMLLDD 362
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 296 -----DYAKUAQYKEDLAAQAAGNAANEADYQAKKAAAYEQELARV-----Q 337
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 363 AENETASILMSGFRQMIHFMTEN-----PDSQAAQQLAAQAAKAAAGDDSAALADA 418
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 338 AANAANK--QAYEQALAAANSKNAQITAENEALQON--AQAKA-----DYAKLAQY 385
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 419 QKALEAA-LKGAGQQQGLNALG-----QIASAAVVSAGVLPLOQ 457
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 386 QKDLAAAQSGNAANEADYQEKLAAYEKELARVQAANAANKAAQAYEQVQOQ 434
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |

RESULT 11
FLJB_SALTY
ID FLJB_SALTY STANDARD; PRT; 505 AA.
AC P52616; P97159;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PHASE-2 FLAGELLIN.
GN FLJB OR H2.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=SL 375;
RX MEDLINE=95325331; PubMed=7541401;
RA Vanegas R.A., Joys T.M.;
RT "Molecular analyses of the phase-2 antigen complex 1,2,.. of
RT Salmonella spp.";
RL J. Bacteriol. 177:3863-3864(1995).
RN [2]
RP SEQUENCE OF 1-37 FROM N.A.
RX MEDLINE=82049491; PubMed=6271461;
RA Silverman M., Zieg J., Mandel G., Simon M.;
RT "Analysis of the functional components of the phase variation
RT system.";
RL Cold Spring Harb. Symp. Quant. Biol. 45:17-26(1981).
RN [3]
RP SEQUENCE OF 482-505 FROM N.A.
RX STRAIN=SJ2353;
RA Mingorance J., Tanaka S., Tomimaga A., Enomoto M.;
RA Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
CC FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -1- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
CC BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLA, TERMED
CC PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -1- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.
CC
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announcement/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U17177; AAC43354.1;
CC EMBL; V01370; CAA24655.1;
CC EMBL; D26168; RAA05156.1;
CC ScyGene; SG10584; fljB.
CC InterPro; IPR001029; Flagellin_C.
CC InterPro; IPR001492; Flagellin_N.
CC Pfam; PF00700; Flagellin_C; 1.
CC Pfam; PF00669; Flagellin_N; 1.
```

```
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 1.
KW Flagella.
FT INIT-MET 0 0 BY SIMILARITY.
FT CONFLICT 37 37 I -> S (IN REF. 2).
SQ SEQUENCE 505 AA; 52404 MW; 3A5CC404AF7AF88B CRC64;

Query Match 6.5%; Score 148; DB 1; Length 505;
Best Local Similarity 21.08; Pred. No. 0.5;
Matches 93; Conservative 59; Mismatches 168; Indels 122; Gaps 20;

QY 23 AQGLEASAAKSAQAQRIAG-----AEAKPKESKTSVERWSILRSVNALMSLADKL 75
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 32 SSGRLINSKADDAAGQAIAANRETANIKGLTQAARNANDGISIAQTTEGALNEINNLRV 91
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 76 GIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTAQ-----TAYDTIFT----- 123
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 92 RELAVQSANSTNSQSDLSIQAEI--TORLNEIDRVSGQTQFNGVKVLAQDNTLTIOYGAN 150
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 124 -----STSLADIAALVSLQ-----DAVTNKTDAATD-----EETAIAAE 159
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 151 DGETIDIDLKQINSQTLGLDSLNVQKAYDKDTAVTTKAYANNCTTLDVSLDAAIAKAA 210
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 160 WETKNADAIKVGAIQITELAKYASDAQILDSLGLKLTDFDLQ-----TALLQS 207
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 211 TGGTNGTASVTGGAV-----KFDADNNKYFTVIGGFTGADAAKNGDYEYVNVATDGTVTLAA 266
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 208 VANNN-----KAELLKEMQDNPV-----PGKTPAIAGSLVDQTDATATQI-----EKD 252
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 267 GATKTTMPAGATTKTEVOELKDTFPAVVSADAKNALIAGG-VDATDANGAELVKMSYTDKN 325
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 253 GNAI-----GDAYFAG--QNASGAVENAKSN-----NSISNIDSAKAAI 289
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 326 GKTEGGYALKAGDKYAAADYDEATGAIK-AKTSYTAADGTTTAANQLGGVD-GKTEV 383
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 290 AT-----AKTQIAEAQKPPDSPILQEAQEQVIAE-----KDLKNIKPA 329
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 384 VTIDGKTYNASKAGHDFKAQPELAAAKTENP-LQKIDAAALQVDAALRSDLGAVQNR 442
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
QY 330 DGSVPNPGTTVGGSKQGGSSI 351
  |||  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 443 FNSAITNLGNTVNNLSEARSRI 464

RESULT 12
YJH8_YEAST
ID YJH8_YEAST STANDARD; PRT; 881 AA.
AC P47033;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCP160-SMC3 INTERGENIC REGION.
GN YJL078C OR J1027.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=S288C;
RX MEDLINE=96093911; PubMed=7483841;
RA Miosga T., Schaaff-Gerstenschlaeger I., Chalwatzis N., Baur A.,
RA Boles E., Fournier C., Schmitt S., Velten C., Wilhelm N.,
RA Zimmermann F.K.;
RT "Sequence analysis of a 33.1 kb fragment from the left arm of
RT Saccharomyces cerevisiae chromosome X, including putative proteins
RT with leucine zippers, a fungal Zn(II)-Cys6 binuclear cluster domain
RT and a putative alpha 2-SCB-alpha 2 binding site.";
RL Yeast 11:681-689(1995).
CC -1- SIMILARITY: BELONGS TO A FAMILY THAT GROUPS MAMMALIAN SCP/TPX1;
CC INSECTS AG3/AG5; FUNGI SC7/SC14 AND PLANTS PR-1.
```

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announcement/> or send an email to license@isb-sib.ch).

EMBL; 249353; CAA89370.1; -
 EMBL; X83502; CAA58492.1; -
 EMBL; X88851; CAA61314.1; -
 HSSP; P04284; ICPE.
 SCD; S0003614; YJL078C.
 InterPro; IPR001283; SCP.
 Pfam; PF00188; SCP; 1.
 SMART; PR00837; V5TPXLIKE.
 PRINTS; SM00198; SCP; 1.
 PROSITE; PS01009; SCP_AG5_PRI_SCT_1; 1.
 PROSITE; PS01010; SCP_AG5_PRI_SCT_2; 1.
 Hypothetical protein.
 FT DOMAIN 166 379 ALA/SER/THR-RICH.
 SEQUENCE 881 AA; 89152 MW; 37DBAC660CA9D12A CRC64;

Query Match 6.5%; Score 148; DB 1; Length 881;
 Best Local Similarity 21.8%; Pred. No. 0.99;
 Matches 100; Conservative 73; Mismatches 199; Indels 86; Gaps 19;

QY 22 SAQGLEASAA--NKSAAQRTAGAEAKPKESKTSVERWSILRSVAVNALMSLADKLGTAS 79
 DB 365 SAQTSSVSASLSSVAADDIOGSTS--KEATSVSEHTSIVTSATNA-AQYATRLGSSS 421
 QY 80 SNSSSVTSRSA-----DVSSTTATPTPPPTSDDYKTKQATQATYDTFT 123
 DB 422 RSSSGAVSSAVSOSVNLNVAVNTDYSVTSVSTA-----HTTKDTATTSVTASEITS 476
 QY 124 STSLADQALVSLQDAVNTKDAATDETAIAAEWETKNADAIVKGAQITELAKYA-- 181
 DB 477 ETAQAS-----SSTKEKINSNAATSSSI-----YSNSASVS-CHGVTYAAEVAIT 520
 QY 182 -SDNQALDLSGLKTSFPLLOTALLOSVA--N--NKAALLKEMODNPVVPCKTPAIA 235
 DB 521 SEQSLATSPATNCSSIVATTLTENSSTTTITATKSTTLATANN-----STRAAT 575
 QY 236 QSLVDQT-DATATQIEKDGNAIGDAYFAGQNASGAVENAKNSNISIDSKAAIATAKT 294
 DB 576 AVTIQPTLDPTDNGASPTDNKHTSTVGSSTGASLDSLRITTSISVSSNTTQLVST--- 632
 QY 295 QIAEAQKFPSPILQEAQVIAQKDL-KNIRPADGSDVPNPGTGVGSKQGGSSIGS 353
 DB 633 --CTSESIDSPSPF--AISTATTESNLITNTITASCSTDSNEPPTSAASTDETAFT 688
 QY 354 IRVS-MLLDDAENETASTLMSGFQIMHENT-----NTVVPASFPSTTTTCLNDDTAFFSIYTE 743
 DB 689 ISTSCSLNGASTQTSELTTSPMT-----NTVVPASFPSTTTTCLNDDTAFFSIYTE 743
 QY 398 AQARAAKAGDSDSAAA-LADAQKALEALGAKAQOOG 434
 DB 744 VNAATIINPGETSSLASDFATSEKPNPTSVKSTSN 781

RESULT 13
 HTR6_HALSA STANDARD; PRT; 778 AA.
 ID HTR6_HALSA
 AC Q48319;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE HALOBACTERIAL TRANSDUCER PROTEIN VI.
 GN HTR4 OR HPVI.
 OS Halobacterium salinarum.
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.

QX NCBI_TaxID=2242;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-S9;
 RX MEDLINE=96275896; PubMed=8674984;
 RA Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
 R Osterheld D.;
 RT "A family of halobacterial transducer proteins.";
 RL FEMS Microbiol Lett. 139:161-188(1996).
 CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
 CC TRANSDUCTION.
 CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).

EMBL; X95590; CAA64843.1; -
 InterPro; IPR000122; Chemotaxis_transducer.
 InterPro; IPR000658; DUF5.
 InterPro; IPR003660; HAMP.
 Pfam; PF00672; DUF5; 1.
 Pfam; PF00015; MCPsignal; 1.
 SMART; SM00304; HAMP; 2.
 SMART; SM00283; MA; 1.
 Transducer; Transmembrane.
 KW DOMAIN 1 26 CYTOPLASMIC (POTENTIAL).
 FT TRANSMEM 27 47 POTENTIAL.
 FT DOMAIN 48 296 EXTRACELLULAR (POTENTIAL).
 FT TRANSMEM 297 317 POTENTIAL.
 FT DOMAIN 318 778 CYTOPLASMIC (POTENTIAL).
 SQ SEQUENCE 778 AA; 82077 MW; 134C7D7F0A3334CD CRC64;

Query Match 6.5%; Score 147; DB 1; Length 778;
 Best Local Similarity 22.9%; Pred. No. 0.96; Indels 58; Gaps 19;
 Matches 97; Conservative 74; Mismatches 184;

QY 20 DLSAQGLEASAAKSAEQAQRTAGAEAKPKESKTSVERWSILRSVAVNALMSL----- 71
 DB 378 DAEEAARSEAAQAADAAQAQ--AEAEAAARESEAAQ-RRLETTAEAFSETMRAYAAGDLTV 435
 QY 72 ---ADKLGIASSN-SSSSTSRSADVSTTATPTPPPTSDDYKTKQATQATYDTFTSTSL 127
 DB 436 RLDRADVQQAAMADIAAFNEMAADMEATADAGR----FADEVATASTDASD-----SA 485
 QY 128 ADIAQALVSLQDAVTNKTAA--TDETAIAAEWETKNADAIVKGAQITELAKYASDNQ 185
 DB 486 AAVEQTGRVSDAVGRIRDRADQDQLEAVASETDEMSATIEEVAASADQVAE-TSORA 544
 QY 186 AILDSLGKLTSLLOTALLOSVA--N--NKAALLKEMODNPVVPCKTPAIAQ-- 237
 DB 545 AALGDGQAAQADAV--AQLEIEDETAATAVDDLEAKMSEIETIVAAITDIAEQTNM 602
 QY 238 LVQDTQATATQIEKDG--AIGDAY--FAGONASGAVE-----NAKSNNSISNIDS 284
 DB 603 LALNANIEAARADQDGGFVAVVADQVADQLEASAKAAEIALVAEVAETETSVAMDR 662
 QY 285 AKAAIATAKTAQIAEAKKFPDPSILQEAQVIAQ--KDLKNIKPADGSDVPNPGTTV 341
 DB 663 IQERVSDGVETVSETRS-----LSEIAGRIAEADTGVQEISNAMDDQAAASVDVTTAV 716
 QY 342 GGSQKQSGSSIGSIRVSMLLDDAENETASILMSGFQIMHENTEN-PDSQAQQLAAQA 400
 DB 717 GDVAALGEETATEAESTA--DAAAEQAQSTLSDVAAQ-----TETLAEHAVALREHAQF 768

401 RAA 403
 769 EVA 771

```

RESULT 14
STF_LAMB
ID STF_LAMB STANDARD; PRT; 774 AA.
AC P03764;
DT 21-JUL-1986 (Rel. 01, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DE SIDE TAIL FIBER PROTEIN.
GN STF.
OS Bacteriophage lambda.
OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
OC Lambda phage group.
OX NCBI_TaxID=10710;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=83189071; PubMed=6221115;
RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
RT "Nucleotide sequence of bacteriophage lambda DNA.";
RL J. Mol. Biol. 162:729-773(1982).
RN [2]
RN IDENTIFICATION AS STF.
RX MEDLINE=92165720; PubMed=1531648;
RA Haggaard-Ljungquist E., Halling C., Calendar R.;
RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence
for horizontal transfer of tail fiber genes among unrelated
bacteriophages.";
RL J. Bacteriol. 174:1462-1477(1992).
RN [3]
RN RECONSTRUCTION OF STF.
RX MEDLINE=93068310; PubMed=1439823;
RA Hendrix R.W., Duda R.L.;
RT "Bacteriophage lambda Paf: not the mother of all lambda phages.";
RL Science 258:1145-1148(1992).
CC -1- MISCELLANEOUS: THE COMMON LABORATORY STRAIN OF BACTERIOPHAGE
CC LAMBDA; LAMBDA PAPA; CARRIES A FRAMESHIFT MUTATION RELATIVE TO UR-
CC LAMBDA, THE ORIGINAL ISOLATE. THE UR-LAMBDA VIRIONS HAVE THIN,
CC JOINTED TAIL FIBERS (SIDE TAIL FIBERS) THAT ARE ABSENT FROM LAMBDA
CC WILD TYPE. RELATIVE TO LAMBDA PAPA, UR-LAMBDA HAS EXPANDED
CC RECEPTOR SPECIFICITY AND ADSORBS TO E. COLI CELLS MORE RAPIDLY.
CC -1- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC CORRECTED IN POSITION TO 396 TO RECREATE THE ORIGINAL STF PROTEIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J02459; AAA96555.1; ALT_FRAME.
DR EMBL; J02459; AAA96557.1; ALT_FRAME.
DR PIR; A04389; QXBP1L.
DR PIR; A04370; QXBP2L.
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR001778; F0A_allergen.
KW Fiber protein.
SQ SEQUENCE 774 AA; 77527 MW; CDD1DF85E919123B CRC64;

Query Match 6.58; Score 146.5; DB 1; Length 774;
Best Local Similarity 20.94; Pred. No. 1;
Matches 90; Conservative 66; Mismatches 209; Indels 65; Gaps 13;

QY 26 LEA5AANKSAQRIAGAEAKPKESKTSVDFWISRLSVAVALMSLADKLGIASSNSSSS 85
DB 110 VEEVARNASVVQASTADAKKSAGDASASAAQVAALVTDATDSARAATSAGQAAS-SAQE 168
QY 86 TSSRADVDSTTATPTPPPTSDDYKTAQATPAYDTIFTSTSLADIQAALVSLQDAVNIK 145

```

```

Db 169 ASSGAEEAASAKATEAEKSAEAAAESSKNAATSAAGAAKTSET-----NAAASQCSAATSA 223
QY 146 DTAATDEETAIAAEWETKNADAIKVGAQITELAKYASDNOAIILDSLGKLTSPFLLQTALL 205
Db 224 TAATKASEAATSA---RDVASKEAAKSETNASSAGRAASATAAENSARAATKS-- 277
QY 206 QSVANNKAAELKEMQDNVPVPGKTPATAQSLVDOTDATATQIEKDGNAIGDAYFAGON 265
Db 278 ---ETNARSSETAAERSASAAADAKTAAGSASTASTKAT-----EAAGSASVASQS 326
QY 266 ASGAVENA-KSNSISNIDSAKAAIA-----TAKTOIAEAKKFPDPSPILOEAEQMVIOA 319
Db 327 KSAEAAAIRAKNSAKRAEDIAVALEADATTRKGIQVLSLSATNSATSETLAATPKAVKV 386
QY 320 EKDLKNIK-PADG---SDVPNPGTTVGGSKQSGSSIGSIRVSMLLDDAENETASILMSGF 375
Db 387 VMDETNRKAPLDSPALTGTPAPTALRG-----NNTQTANTAFVIAAI 430
QY 376 ROMIHMENTENPDQSAQAQAAKAAAGDDSAALADAKKALEAALGRAGQOOGI 435
Db 431 ADVI---DASPDALNTLNELAA-----ALGND-----PDFATTMTNALAGKQPKINAT 474
QY 436 LNALQTIASA 445
Db 475 LTALAGLSTA 484

RESULT 15
SYEP_DROME
ID SYEP_DROME STANDARD; PRT; 1714 AA.
AC P28668; Q9VCF5;
DT 01-DEC-1992 (Rel. 24, Created)
DT 01-DEC-1992 (Rel. 24, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE BIFUNCTIONAL AMINOACYL-TRNA SYNTHETASE [INCLUDES: GLUTAMYL-TRNA
DE SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE); PROLYL-TRNA
GN AATS-GLUPRO OR CG5394.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Phrygotea; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OX NCBI_TaxID=7227;
RN [1]
RN SEQUENCE FROM N.A.
RX MEDLINE=92057547; PubMed=1756734;
RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;
RT "A component of the multisynthetase complex is a multifunctional
aminoacyl-trna synthetase.";
RL ENBO J. 10:4267-4277(1991).
RN [2]
RN SEQUENCE FROM N.A.
RX STRAIN=OREGON-R;
RX MEDLINE=97217441; PubMed=9063462;
RA Cerini C., Semeriva M., Gratecos D.;
RT "Evolution of the aminoacyl-trna synthetase family and the
organization of the Drosophila glutamyl-prolyl-trna synthetase gene.
Intron/exon structure of the gene, control of expression of the two
mRNAs, selective advantage of the multi-enzyme complex.";
RL Eur. J. Biochem. 244:176-185(1997).
RN [3]
RN SEQUENCE FROM N.A.
RX STRAIN=BERKELEY;
RX MEDLINE=20196006; PubMed=10731132;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Ananatis P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Vandeil M.D., Zhang Q., Chen L.X.,
RA Brannon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA Abril J.F., Agapayni A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,

```

RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
 RA Foster C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Laslo P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Svirskas R., Tector R., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissenbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RA "The genome sequence of *Drosophila melanogaster*.";
 RT Science 287:2185-2195(2000).

CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +

CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).

CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(PRO) = AMP +

CC PYROPHOSPHATE + L-PROLYL-TRNA(PRO).

CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS

CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE

CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,

CC ARGINYL, AND ASPARYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY

CC PROTEINS, P18, P48 AND P43.

CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I

CC AMINOACYL-TRNA SYNTHETASE FAMILY.

CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II

CC AMINOACYL-TRNA SYNTHETASE FAMILY.

CC -1- SIMILARITY: CONTAINS 6 "WHEP-TRS" DOMAINS.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: M74104; AAA28594.1; -

CC EMBL: U59923; AAC47469.1; -

CC EMBL: AE003745; AAF56211.1; -

CC PIR: S18644; S18644.

CC HSP: P00962; IORU.

DR Flybase: FBgn005674; Aats-glupro.

DR InterPro: IPR002106; AA_trna_ligase_II.

DR InterPro: IPR000738; WHEP-TRS.

DR InterPro: IPR000924; trna-synt_1c.

DR InterPro: IPR002314; trna-synt_2b.

DR InterPro: IPR001412; trna-synt_1.

DR InterPro: IPR002316; trna-synt_pro.

DR Pfam: PF00749; trna-synt_1c; 1.

DR Pfam: PF00587; trna-synt_2b; 1.

DR Pfam: PF00458; WHEP-TRS; 6.

DR PRINTS: PR00987; TRNASYNTHGLU.

DR PRINTS: PR01046; TRNASYNTHPRO.

DR PROSITE: PS00178; AA_TRNA_LIGASE_I; 1.

DR PROSITE: PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.

DR PROSITE: PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.

DR PROSITE: PS00762; WHEP-TRS; 6.

KW Aminoacyl-trna synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Multifunctional enzyme; Repeat.

FT DOMAIN 170 754 GLUTAMYL-TRNA SYNTHETASE.

FT DOMAIN 755 800 WHEP-TRS 1.

FT DOMAIN 827 872 WHEP-TRS 2.

FT DOMAIN 901 946 WHEP-TRS 3.

FT DOMAIN 980 1025 WHEP-TRS 4.

FT DOMAIN 1055 1100 WHEP-TRS 5.

FT DOMAIN 1129 1173 WHEP-TRS 6.

FT DOMAIN 1174 1180 POLY-GLY.

FT DOMAIN 1207 1714 PROLYL-TRNA SYNTHETASE.

FT SITE 209 220 "HIGH" REGION.

FT SITE 438 442 "KMSKS" REGION.

FT BINDING 441 441 ATP (BY SIMILARITY).

FT CONFLICT 102 106 TSPLP -> DKSIA (IN REF. 3).

FT CONFLICT 233 234 VC -> AF (IN REF. 3).

FT CONFLICT 341 345 NTACA -> KYCVR (IN REF. 3).

FT CONFLICT 583 583 K -> R (IN REF. 3).

FT CONFLICT 692 692 L -> A (IN REF. 3).

FT CONFLICT 753 753 T -> S (IN REF. 3).

FT CONFLICT 802 802 T -> S (IN REF. 3).

FT CONFLICT 873 873 P -> T (IN REF. 3).

FT CONFLICT 887 887 G -> V (IN REF. 3).

FT CONFLICT 1201 1201 P -> PA (IN REF. 3).

FT CONFLICT 1461 1461 MISSING (IN REF. 3).

FT CONFLICT 1587 1587 G -> V (IN REF. 3).

FT SEQUENCE 1714 AA; 189197 MW; 6FE8C58045E48A8C CRC64;

Query Match 6.5%; Score 146; DB 1; Length 1714;

Best Local Similarity 19.8%; Pred. No. 2.8;

Matches 101; Conservative 74; Mismatches 192; Indels 142; Gaps 20;

Oy 7 PGPI-----DETERTPADLSAAGLEASAAKSAEAQRIAG-----A 43

Db 701 PSPVILFIPGHTKDVPTSLGLKVNAPDAKATKAKSSVSSGGQASELDSQITQGDVLR 760

Oy 44 EAKPKESKTSVERWSILRSVAVNMLSLADKLGTASSSSSSSSSSSADVDSTAT-APTTP 102

Db 761 DLKSKKAADQID-----VAVKKLLAL-----KADYKSAATGDKWPGQTSATSAVPV 807

Oy 103 PPTPSDDYKTAQATAYDTFTSTSLADIAQALVSLQDAVNIKDTAATDEETAIA----- 157

Db 808 AASSS-----SANDAV-----SYNASIVKOGDLVRDLKGRKASKPEIDAIAVKTLL 852

Oy 158 ---AEWETKNADAIVKGAQITELAKYASDNOAI---LDSLGKLTSPDLLQTLALQSVANN 212

Db 853 ELKAGYKTLTGQDKWPGVPTAAPSASAAPSVGVGNDVSAQILS-----QITAGOD 903

Oy 213 KAAELLKEMQDNVVPVPGKTPAIAOSLVDPDTATQIEKDGNAIGDAYFAGONASGAVEN 272

Db 904 KVRELKSAKADKATVDAAVKTLLSLKADYKAATGSD-WKPGTT-----APAPAAAPVKV 956

Oy 273 AKSN--SISNIDSAAKATATAKTATAEAKKFPDPIQEAQOMVIAQEKDLKNI-- 326

Db 957 KOEKNPDPASVLTNTLLNKIAQGGDKIRQLKSAKSEKSLVEAEVKLLALKTDYKSLTG 1016

Oy 327 ---RPADGSDVFNFGTT-----VGGSKQGGSGISGSI----- 354

Db 1017 QEWKP--GTVAAPTPTVNVIDLTTGG--DSGSDVGVSVLSKIAQGGDKIRKLKSEKAANKVI 1072

Oy 355 --RVSMLL-----DDAENETASILMSGFQMIHMTENTENPOSOAAQOELA 397

Db 1073 DPEVKTLALKGEYKTLGKDKWTDPKSEPAVV-----KKEASPVSMASPAKDELTOEIN 1127

Oy 398 AQAARAAGDDSAALADAKALEAAL 426

Db 1128 AQGEVRAAKGNKAAKEVIDAEVAKLLAL 1156

Search completed: February 7, 2002, 21:42:45

Job time: 599 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:15 ; Search time 172 Seconds
(without alignments)
393.745 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MNPITGPGPIDETERTPPAD.....SAAVAGVLPQQLWIRA 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues

Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_17.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match	Length	DB	ID	Description
1	2214	97.9	715	2	Q46166	Q46166	chlamydia p
2	2188	96.8	651	2	Q927H7	Q927H7	chlamydia p
3	451.5	20.0	647	2	Q84627	Q84627	chlamydia t
4	192	8.5	1327	2	Q9X7M2	Q9X7M2	staphylococ
5	192	8.5	2478	2	Q9RL69	Q9RL69	staphylococ
6	190	8.4	2478	2	Q9LCH2	Q9LCH2	staphylococ
7	182.5	8.1	2055	2	Q85472	Q85472	staphylococ
8	182	8.0	2481	2	Q99QR6	Q99QR6	staphylococ
9	180	8.0	2271	2	Q99QY4	Q99QY4	staphylococ
10	177.5	7.9	1822	2	Q07290	Q07290	staphylococ
11	177	7.8	697	5	Q9NDJ0	Q9NDJ0	plasmodium
12	174	7.7	1545	7	Q9RD01	Q9RD01	staphylococ
13	173	7.7	2016	5	Q9BIT0	Q9BIT0	staphylococ
14	171	7.6	627	1	Q9HS86	Q9HS86	staphylococ
15	171	7.6	993	2	Q99QZ5	Q99QZ5	staphylococ
16	171	7.6	1795	2	Q9LJC9	Q9LJC9	staphylococ
17	166.5	7.4	1128	5	Q26947	Q26947	trypanosoma
18	166.5	7.4	2232	5	P91365	P91365	caenorhabdi
19	164	7.3	6713	2	Q99U54	Q99U54	staphylococ

20	161.5	7.1	956	2	Q9LON7	Q9LON7	streptomyce
21	161.5	7.1	2045	2	Q9A0X5	Q9A0X5	streptococc
22	161.5	7.1	3381	2	Q9XK33	Q9XK33	streptococc
23	161	7.1	1038	10	Q9AS09	Q9AS09	oryza sativ
24	160.5	7.1	544	1	P71409	P71409	halobacteri
25	160.5	7.1	2178	2	Q9KWR3	Q9KWR3	streptococc
26	159	7.0	1579	11	Q9WMP1	Q9WMP1	mus musculu
27	159	7.0	1684	11	Q9WTQ5	Q9WTQ5	mus musculu
28	158.5	7.0	1156	2	Q9Z5A4	Q9Z5A4	streptomyce
29	158	7.0	1365	2	Q9Z5Z5	Q9Z5Z5	mycoplasma
30	158	7.0	2186	2	Q99TB0	Q99TB0	staphylococ
31	157	6.9	1344	2	Q9Z5A4	Q9Z5A4	mycoplasma
32	156.5	6.9	641	2	Q9KSS7	Q9KSS7	vibrio chol
33	156	6.9	990	13	Q91803	Q91803	xenopus lae
34	156	6.9	6677	5	Q9A435	Q9A435	caenorhabdi
35	155	6.9	845	5	Q9Y1P8	Q9Y1P8	plasmodium
36	155	6.9	892	3	Q9P3P5	Q9P3P5	neurospora
37	154.5	6.8	1262	5	Q20684	Q20684	caenorhabdi
38	154	6.8	1041	10	Q9ASA4	Q9ASA4	oryza sativ
39	154	6.8	1814	5	Q9B1S9	Q9B1S9	plectreury
40	153.5	6.8	969	5	Q9NDI9	Q9NDI9	plasmodium
41	153.5	6.8	993	10	Q9ASL3	Q9ASL3	oryza sativ
42	153	6.8	1569	2	Q94183	Q94183	streptococc
43	153	6.8	2614	5	Q97054	Q97054	dictyosteli
44	152.5	6.7	653	2	Q9X6N1	Q9X6N1	rhizobium l
45	152.5	6.7	962	2	Q9RIE3	Q9RIE3	yersinia pe

ALIGNMENTS

RESULT 1

Q46166 PRELIMINARY; PRT; 715 AA.
AC Q46166;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE 76 KDA PROTEIN.
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94156481; PubMed=7509320;
RA Perez-Melgosa M., Kuo C.C., Campbell L.;
RT "Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kilodalton protein containing a species-specific epitope";
RL Infect. Immun. 62:880-886(1994).
DR EMBL; L23921; AAA23117.1;
SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 97.9%; Score 2214; DB 2; Length 715;
Best Local Similarity 98.9%; Pred. NO. 3.9e-115;
Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

Qy	1	MVNPITGPGPIDETERTPPADLSAOGLEASAAKSAEAOIAGAEAKPKESKTSVRSWI	60
Db	257	LWNPITGPGPIDETERTPPADLSAOGLEASAAKSAEAOIAGAEAKPKESKTSVRSWI	316
Qy	61	LRSAVNALMSLADKLGIASSSSSSSSRSADVDSSTATATPTPPPTSDYKTAQATYDT	120
Db	317	LRSAVNALMSLADKLGIASSSSSSSSRSADVDSSTATATPTPPPTSDYKTAQATYDT	376
Qy	121	IFTSTSLADIQAALVSLQDAVNTNIKDTATDEETAAEWEETKNADAKVGAQITELAKY	180
Db	377	IFTSTSLADIQAALVSLQDAVNTNIKDTATDEETAAEWEETKNADAKVGAQITELAKY	436
Qy	181	ASDNOAILDLSGLKTSFLLQATALLQSVANNKAAELKEMODNPVPGKTPAIAGSLVD	240
Db	437	ASDNOAILDLSGLKTSFLLQATALLQSVANNKAAELKEMODNPVPGKTPAIAGSLVD	496

Query Match 96.8%; Score 2188; DB 2; Length 651;
Best Local Similarity 98.9%; Pred. No. 9.7e-114;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
QY 1 MYNPICGPIDETPTPADLSAAGLPSAANKSAEQAQIAGAAKPKESKTDTSVERWSI 60
|||||

	Query Match	20.0%	Score 451.5;	DB 2;	Length 647;
	Best Local Similarity 29.6%;		pred. No. 1.6e-17;		
	Matches 138;	Conservative 85;	Mismatches 192;	Indels 51;	Gaps 14;
Qy	2	VNPICPGPIDETERTPPADLSAOGLEASAAKSAEAORIAGAEAKPKESKTDVSVEEWSIL	61		
Db	15	MNPINGOI-----ASNTEKSTKESEA-----SPSASSSVSSWSFL	52		
Qy	62	RSAYNALMSLADKLGIASSNSSSTSRSDADVTATATPTPPPTSDDY-----KTQAQATAY	118		
Db	53	SSAKHALISLRD--AILNKNSSPTDSLQ-OLEASTSTS-TVTRVAARDYNEAKSNFDTAK	108		
Qy	119	DTIFTSLSADIQAALVSLQDAVNIKDTAATDEETAIAAEWEKKNADAIVKGAQLTELA	178		
Db	109	SGLENAATTAEYETKMADLMAALQOEMRLAKOKAEVTRIKALQKQEVID--KLNQIV	165		

[illegible]

```

RA Mengin-Lecreulx D., Tomas A.;
RT The femR315 gene from Staphylococcus aureus, the interruption of
RT which results in reduced methicillin resistance, encodes a
RT phosphoglucosamine mutase.;
RL J. Bacteriol. 179:5321-5325(1997).
[3]
RN
RP SEQUENCE FROM N.A.
RT STRAIN=COL;
RX MEDLINE=99265121; PubMed=10332717;
RA Wu S., de Lencastre H.;
RT "Mrp-a new auxiliary gene essential for optimal expression of
RT methicillin resistance in Staphylococcus aureus.";
RT Microb. Drug Resist. 5:9-18(1999).
RD EMBL; Y09927; CAB55329.1;
SQ SEQUENCE 2478 AA; 263029 MW; 6B9859A02D023C74 CRC64;

Query Match. 8.5%; Score 192; DB 2; Length 2478;
Best Local Similarity 23.9%; Pred. No. 0.02;
Matches 133; Conservative 86; Mismatches 194; Indels 144; Gaps

QY 4 PICGPIDE-----TERTPADLSAQGLEASANKSA-EAORIACAENAKPKESKTDs--- 54
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 780 PLKPDPTTNEEVAEATERINAARVS--GVKATEATTAQDLERVKNEEISKETITDSTQT 837
| : | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 55 -----VERWSILRSVAVALMS-----LADKLG-----IASSNSSSSTS 87

```

```

Db 838 KMDAYNEVQQAATARKAQAATVSNATNEEVAEADAADAAQKQGLHDIQVVKSKQEVADT 897
QY 88 RSADVSTTA-TAPTPTPTSDDYKTAQATAYDT-----IFTSTSLADIQAALVSL--- 137
Db 898 KSKVLDKINAQIQAQKVPKPAAD---TEVENAYNTRKQEIQNSASTTEEKQAAAYTELDTK 954
QY 138 -QDAVTNIKDTAATD-----BETAIAAEWETKNADAIKVGQAQITELAKYASDNOAILDS 190
Db 955 KQEAARTNL-DAANTNSDVTTAKDNSIAAINOVQAAATTKKSDAK-AEIAQKASERKTAIEA 1012
QY 191 LGKLT-----SPDLLOALLOSVAANNKAAELLKEMQDNPVVPKGT----- 231
Db 1013 MNDSTTEEQAAKDKVQAVVTANADIDNAA-----ANNVDVNKATTNEATIAATPDA 1066
QY 232 ---PALAQSLVDQDATATQIEKDN-----AIGDAYFAGONASGA 269
Db 1067 NVKPAQAQIADKVOAQETAI--DGNNGSTTEEKAAAKQOVQTEKTTADAADAIDAAHTNAE 1124
QY 270 VENAKSNNSISNI-----DSAKAAIAT-----AKTOIAEAKKFPDPSPILOEA--E 313
Db 1125 VEAAK-KAAIAKIEAIQAPATTTKDNKAELATKANERKTAIAQOTDITABEIAAANADVD 1183
QY 314 QMWIOAEKDLNKPADG-SDVPNPGTTVGSGKQOQSSIGSIRVSMILLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVDQAKTTGENSIDVPTVYKAT-----ARNEITAILN 1235
QY 373 SGFRQMIHMFNTENPDQAAQOELAAQARAAGAAGDSAAA-----ALADAQKALEAA 425
Db 1236 NKLEIQATPDATDEEKQAAAE--ANTENGKANQAISAAATNAQVDEAKANAEEAINAV 1293
QY 426 LGKAGOOQOILNALGOI 442
Db 1294 TPKVVKQAQKDEIDQL 1310

```

```

RESULT 6
ID Q9LCH2 PRELIMINARY; PRT; 2478 AA.
AC Q9LCH2;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE FMTB.
GN FMTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RA Komatsuzawa H.;
RT "Staphylococcus aureus gene for affecting the methicillin resistance."
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB025716; BAA93438.1;
SQ SEQUENCE 2478 AA; 262993 MW; 1C118EBE0DB03B34 CRC64;

```

```

Query Match 8.4%; Score 190; DB 2; Length 2478;
Best Local Similarity 23.9%; Pred. No. 0.026;
Matches 133; Conservative 85; Mismatches 195; Indels 144; Gaps 27;

QY 4 PTGPGPIDE-----TERTPPADLSAQGLEASAANKSA-EAORIAGAACKPKESKTDSDS--- 54
Db 780 PLNPOTTEVEAEALERNNAKVS--GVKAIEATTTAQLERKNEEISKENITDSTQOT 837
QY 55 -----VERWSILRSAYNALMS-----LADKLG-----IASSNNSSSSTS 87
Db 838 KMDAYNEVQQAATARKAQAATVSNATNEEVAEADAADAAQKQGLHDIQVVKSKQEVADT 897
QY 88 RSADVSTTA-TAPTPTPTSDDYKTAQATAYDT-----IFTSTSLADIQAALVSL--- 137

```

```

Db 898 KSKVLDKINAQIQAQKVPKPAAD---TEVENAYNTRKQEIQNSASTTEEKQAAAYTELDTK 954
QY 138 -QDAVTNIKDTAATD-----BETAIAAEWETKNADAIKVGQAQITELAKYASDNOAILDS 190
Db 955 KQEAARTNL-DAANTNSDVTTAKDNSIAAINOVQAAATTKKSDAK-AEIAQKASERKTAIEA 1012
QY 191 LGKLT-----SPDLLOALLOSVAANNKAAELLKEMQDNPVVPKGT----- 231
Db 1013 MNDSTTEEQAAKDKVQAVVTANADIDNAA-----ANNVDVNKATTNEATIAATPDA 1066
QY 232 ---PALAQSLVDQDATATQIEKDN-----AIGDAYFAGONASGA 269
Db 1067 NVKPAQAQIADKVOAQETAI--DGNNGSTTEEKAAAKQOVQTEKTTADAADAIDAAHTNAE 1124
QY 270 VENAKSNNSISNI-----DSAKAAIAT-----AKTOIAEAKKFPDPSPILOEA--E 313
Db 1125 VEAAK-KAAIAKIEAIQAPATTTKDNKAELATKANERKTAIAQOTDITABEIAAANADVD 1183
QY 314 QMWIOAEKDLNKPADG-SDVPNPGTTVGSGKQOQSSIGSIRVSMILLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVDQAKTTGENSIDVPTVYKAT-----ARNEITAILN 1235
QY 373 SGFRQMIHMFNTENPDQAAQOELAAQARAAGAAGDSAAA-----ALADAQKALEAA 425
Db 1236 NKLEIQATPDATDEEKQAAAE--ANTENGKANQAISAAATNAQVDEAKANAEEAINAV 1293
QY 426 LGKAGOOQOILNALGOI 442
Db 1294 TPKVVKQAQKDEIDQL 1310

```

```

RESULT 7
ID Q85472 PRELIMINARY; PRT; 2055 AA.
AC Q85472;
DT 01-NOV-1998 (TrEMBLrel. 08, Created)
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
DE EXTRACELLULAR MATRIX BINDING PROTEIN (FRAGMENT).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Abiotrophia.
OX NCBI_TaxID=46125;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NVS-47;
RA Manganelli R., van de Riijn I.;
RT "Cloning and Characterization of emb, a Gene Encoding the Major Adhesin of Streptococcus defectivus."
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF067776; AAD03320.1;
FT NON_TER 2055
SQ SEQUENCE 2055 AA; 215640 MW; 9699C11DDDE93E2FD CRC64;

```

```

Query Match 8.1%; Score 182.5; DB 2; Length 2055;
Best Local Similarity 22.9%; Pred. No. 0.054;
Matches 109; Conservative 76; Mismatches 189; Indels 101; Gaps 21;

QY 19 ADLSAQGLEASAANKSAEQAQIAGAEA-----KPKESKTDSDVERWSILRSAYNALMSIAD 73
Db 1371 ATKAKNAIDAATSDNETAKQNEGTQAINAVPQTPPKATD-----AKNAVTOAAD 1420
QY 74 KLGTASSNNSST-----SRSAVDVSTTATPTPPPTSDDYKTAQATAYDTIFTSLSA 128
Db 1421 RKKDAIENDPNLTREEKVAQAKVDA-----EAKKAKDAIDAATSA 1462
QY 129 DIQAALVSLQDAVTNIKD--TAATDEETAIAAEWETKNADAIVGAQITELAKYASDNOQA 186
Db 1463 DVTAKQNEGTKAINDVPTPTTAKTDAKNAVTOAADAKK-DAIEKDPNLTREEKDA--KA 1519

```

```

996  MNDSTTEEQAAKDKVDQAVVTANADINATANTDVONAKTTNEATIAATPDANKPAA 1055
235  AQLVVDQTDATQIEKDGA-----IGDAYFAGQNASCAVENAKSN 277
1056  KQAIADKVQQAETAIDANNNGSTTEEKAAKQVQTEKTAADAAIDAAHSNVEEAAK-NA 1114
278  SISNI-----DSAKAATAT---AKTQIAEAOKKFPDSPILQEA--EQWVIOAEK 321
1115  EIAKIEAIGAPATTTKDNAKQATATKANERKTAIAQTQDITAEETIAANADVNDNAVTOAN- 1173
322  DUANKIPADG-SDVPPNGTTVGSGQQSGSIRVSMLLDDAENETASTILMSFGROMIH 380
1174  --SNIEAANSQNDVDAQKTGTSETSIDQVTPTVNKKAT-----ARNEITAILNNKLQEIQA 1226
381  MFNTEPDSQAQOELAAQARAKAAGDDSA--ALADAQKALEALGKAGQOQ 433
1227  TPDATDEEKQADAE--ANTENGKQAISAATTTNAQVDEKAKNAEAAINAVTPKVVKKQ 1284
434  GILNALGQI 442
1285  AAKDEIDQL 1293

```

RESULT	9	
Q990Y4	PRELIMINARY;	PRT; 2271 AA.
AC	Q990Y4;	
DT	01-JUN-2001 (TRENBLrel. 17, Created)	
DT	01-JUN-2001 (TRENBLrel. 17, Last sequence update)	
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)	
DE	SA2447 PROTEIN.	
DE	SA2447.	
GN	Staphylococcus aureus subsp. aureus N315.	
OS	Bacteria; Firmicutes; Bacillus/Clostridium group;	
OC	Bacilli/Staphylococcus group; Staphylococcus.	
OC	NCBI_TaxID=158879;	
RP	[1]	
RP	SEQUENCE FROM N.A.	
RA	Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,	
RA	Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,	
RA	Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,	
RA	Takabashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,	
RA	Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,	
RA	Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,	
RA	Ogasawara N., Hayashi H., Hiramatsu K.;	
RT	"Whole genome sequencing of methicillin-resistant Staphylococcus	
RT	aureus."	
RL	Lancet 357:1225-1240(2001).	
DR	ENBL; AF003136; BAB43752.1; -.	
DR	Complete proteome.	
Q990Y4	SEQUENCE 2271 AA; 227847 MW; 7C2A7040D6C8289D CRC64;	

Query Match	8.0%;	Score 180;	DB 2;	Length 2271;
Best Local Similarity	18.9%;	Pred. No. 0.084;		
Matches	79;	Conservative 96;	Mismatches 187;	Indels 56; Gaps
QY	22	SAOGLSEASAAKSAFAQRTAGAEAKPKESKTDTSVERVILRSVNAWLSADKLGIASSN	81	
Db	1516	SASASSEDSSSVTSLDSTSSASMQSSSESQSTSSASLSDSLSTSTSNRMST-----IASLS	1570	
QY	82	SSSSTGRSADVGGTATATPPPTTDDYKTKQAQT-AYDTTFTSLADIQAALVSLQDA	140	
Db	1571	TSVSTSESGTSESTSESDSTSTSLDSQSTSRSTSSASGSASTSTSTSDRSTSASTS--	1628	
QY	141	VTNIKDTAATDEETATAAEWETKNADAKVGAQITELAKYASDNOAILDLSLGLTSLFDL-	199	
Db	1629	-TSMRTSTSDSQSMLSLSTSTSTSMDS-----TSLSDSVSDSTSS--DSTSASTSGMS	1678	
QY	200	LOTALLQSVANNKKAELKEMODPNVPGTKPTAQLSVQDTDATATQIEKQGNATGDA	259	
Db	1679	VSTLSLSDSTSTSTSAEVNMSAISD-----SQSMSESVNDSSEVSNSEDSKSMGSS	1732	

QY 260 YFAGONASGAV-----ENAKSNNSISNIDSAKAAIATAKTAQIAEAQKFPDPSILQEA 312
 Db 1733 TSVSDSGSLSVSTSLRKESVSESSLSGQSMDSVSTSDSSLSVSTSLRSESSES 1792
 QY 313 EQMVIQAEKDLNKPAGDVPNPGTGVGSKQGGSSIG-----SIRVSMILDDA 363
 Db 1793 D-----SLSDSKSTSGS-----TSTSTSGSLSTSTSLSGSESVSESTSLSDSISMSDS 1840
 QY 364 ENETASILMSGFROMIHMFNTENPDQAQOELAAQARAACAAGDDSAALADAQA 421
 Db 1841 TSTSDSLSGSLSGISLSTSLSTSDLSKSLSS-----SQSMGSESTSTSDSQSS 1895
 RESULT 10
 Q07290 PRELIMINARY; PRT; 1822 AA.
 AC Q07290;
 DT 01-NOV-1996 (Tremblrel. 01, Created)
 DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE EF PROTEIN.
 GN EPP*.
 OS Streptococcus suis.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1307;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1890;
 RX MEDLINE=93328280; PubMed=8333363;
 RA Smith H.E., Reek F.H., Vecht U., Gielkens A.L.J., Smits M.A.;
 RT "Repeats in an extracellular protein of weakly pathogenic strains of
 RT Streptococcus suis type 2 are absent in pathogenic strains.";
 RL Infect. Immun. 61:3318-3326(1993).
 DR EMBL; X71880; CAA50714.1;
 SQ SEQUENCE 1822 AA; 192631 MW; 3838960C77641D7D CRC64;

Query Match 7.98; Score 177.5; DB 2; Length 1822;
 Best Local Similarity 23.68; Pred. No. 0.088;
 Matches 116; Conservative 77; Mismatches 195; Indels 103; Gaps 21;

QY 13 TERTPPADLSAQGLEASAAKSAEARIAGAEAKPKES-KTDSVVERWSIL-RSAVNALMS 70
 Db 1201 TEKAKAAELAGEKSLDTGKEARDAVELAKDKELAKEAIRTEEEEAATKIVEKLAEDTRKA 1260
 QY 71 LADKLGIASSNSSSTSSAD-VSTTATATPTPTPTSDDYKTQATAYDTFTSTSLAD 129
 Db 1261 IEDPNLSDEKQAEIKKLTDAVAKTLATI-----RDNADKRTQEAERKA-----QALAD 1309
 QY 130 IQAALVSLQDAVTNIDKTAATDEETAIA--AEWETKNADAIVKGAQITELAKYASDNOAI 187
 Db 1310 LEKA-----KETQKTADKAIDRLTILYKDGLEATKQDAKNKIADKAAAEKATASPN 1364
 QY 188 LDSLGLKLSFDLLOLQSVANNKA-----AELKEMQDNVVPKGTPTAIQAQSLVD 240
 Db 1365 LTDAEKKFTD-----AVDAEAKANDAIASAATSPADVQKE-EDAGVA-----ATAEDVLD 1414
 QY 241 QTDATA-TQIEKDNGAIGDAYFAGONASGA-----VENAKSNNSISNIDSAKAAI 289
 Db 1415 AAKQDAKNKIADKAAAEKATGNSPNLTDAEKKTTFDVADEAVAKANDAIS----- 1465
 QY 290 ATAKTQIAEAQKFPDPSILQAEQMWIQAEDKLNKPAGDSV-----PNPGTT--- 340
 Db 1466 --AATSPADVQKE-EDAGVAATAEDVLDAAKQDAKN-KIAKESDAKSAIDANPNLTDAE 1521
 QY 341 -----VGGSKQGGSSIGSIRVSM-----LDDAENETASILM 372
 Db 1522 KESAKKAVDADAKAATDAIDASTSPVEAQSAEDKGVGSIADVLDAKQDAKNKIADKAA 1581
 QY 373 SGFRQMIHMFNTENPDQAQOELAAQARAACAAGDDSAALADAQAQALEAALGKAGQ 432
 Db 1582 AAKEAIDANPNLSDAEKAERKAVDADAKATTDAID--ASTSPVEAQSAEDKGVGSI--R 1637

QY 433 QGILNALGOIA 443
 Db 1638 QDVLDAAKODA 1648
 RESULT 11
 Q9NDJ0 PRELIMINARY; PRT; 697 AA.
 AC Q9NDJ0;
 DT 01-OCT-2000 (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 3B.
 GN MSP-3B.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BELEM;
 RA Galinski M.R., Al-Khedery B., Ingravallo P., Corredor-Medina C.,
 RA Barnwell J.W.;
 RT "Plasmodium vivax merozoite surface proteins-3 beta and -3 gamma share
 RT structural similarities with Plasmodium vivax MSP-3 alpha and define a
 RT new gene family in Plasmodium.";
 RL Submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF099662; AAF78287.1;
 DR InterPro: IPR000122; Chemotaxis_transducer.
 SQ SEQUENCE 697 AA; 75685 MW; 0AF2AE9801A956A7 CRC64;

Query Match 7.88; Score 177; DB 5; Length 697;
 Best Local Similarity 20.9%; Pred. No. 0.029;
 Matches 118; Conservative 74; Mismatches 186; Indels 186; Gaps 19;

QY 1 MNPVIGPGPIDETERTPPA-----DLAQGLEASAAKSAEARI 40
 Db 53 IVNPDGSDVDTEDEGGEALGQSGPEKSAEPKVAHQAEVNEKSLSKAQAQAAEKA 112
 QY 41 AGAEAKPKESKTDSEVERWSI-----LRSAVN 66
 Db 113 AKAAESAKKNTLDALKVNVPTLNNVKKFAESAATAEQKQENIATAEKKVAEANGEV 172
 QY 67 ALMSLADKLGIASSNS-----SSSTSRSDVDSTTA----- 97
 Db 173 ELQKLKDEVDKAAKAKQQLKAEIAEHAVKAQVAKTEAEKAKQDATTAKEVAIKETGTS 232
 QY 98 -----TAPTPPTSDDYKTQATAYDTFTSTSLA-----DIAAALVSLQDAVT 142
 Db 233 KSENVTKAIDMAKKEEETKNQASIASENADKAAKAAQAEVKEIKDENKEISQLENEIT 292
 QY 143 NIKTATATDETAIAEWETKNADAIVKGAQI-----TELAKYASDNOALDLSGLKITS 196
 Db 293 KLGIDILNTVKELASNAEDASNAKKEKMWKQIAEVAEKAKEAEKAEANFLAEKAK--- 349
 QY 197 FDLQTLALQSVANNKAAELLKEMQDNVVPKGTPTAIQAQSLVDOTDATAFQIEKDNGAI 256
 Db 350 ---QTA-EKIATKSTKITE-----EVRRATEFAKTAGDATTQAATE-----AA 391
 QY 257 GDATFAGONASGAVENAKSNNSISNIDSAKAAIATAKTO---IAEAQKFPDPSILQE-- 311
 Db 392 GDVSSSEKQKQVLLSEIK-QKAESALQASKDAI-KAKTEAENFLEIAKEVKEPAEAAKDEA 449
 QY 312 -----AEQMVIOAEKDLNKPAGDSVVPNPGTTVGGSKQGGSSIGSIRVSMILDDAEN 365
 Db 450 QKATSADEAKTEALKIAEYVKNKSDAS-----EN 478
 QY 366 ETATILMSGFRQMIHMFNTE-NPDSQAQOELAAQARAACAAGDDSAALADAQAQALEA 424
 Db 479 EKKKI-----ETEANATAGEAQKAAFAKAAKAAKADAKDTNEAVTTLAVAKEKVEK 526
 QY 425 ALGRAGQOQGGILNALGQIATAA 448


```
Db 1689 AAAASASS 1696
RESULT 14
Q9HS86 PRELIMINARY; PRT; 627 AA.
AC Q9HS86;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HTR14 TRANSDUCER.
GN HTR14 OR VNG0355G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OC NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
RA Leithauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
DR EMBL; AE004994; AAG18922.1;
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR000658; DUF5.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; DUF5; 1.
DR Pfam; PF00015; MCPsSignal; 1.
DR PRINTS; PR00260; CHEMTRNSDUCR.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
KW Complete proteome.
SQ SEQUENCE 627 AA; 65655 MW; E5F2867FA6CA2A75 CRC64;

Query Match 7.6%; Score 171; DB 1; Length 627;
Best Local Similarity 23.4%; Pred. No. 0.054;
Matches 120; Conservative 87; Mismatches 184; Indels 122; Gaps 27;
QY 17 PPADLSAQGLEASANKSEAEQRIAGAEAKPK-----ESKTDVSVERWS-ILRSVAVNALMSL 71
DB 145 PALDESVPVGFGESEITEMADSLSEAYTALEEDKTALEHQQAELERQSEOLRALVDALSEA 204
QY 72 ADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDDYKTAQ-----QTAYDTIFTS 124
DB 205 TD-----AARAGDL---TATVDAAALDVTDDHRAAEDFNOLLEATDISI 249
QY 125 TSLADIQAAL-----VSLQDAVNIKDTA--ATDEETAIAAEWETKNADA 167
DB 250 QSFSDAVLAVSRITDERVDAVDRSAVSEVTEIADGANQNTNOLNIAAEMDTVSATV 309
QY 168 IKVCAQITELAKYASDQAIL---DSLGLK-TSPDLLQALLOSANNKAAELLKEMO 222
DB 310 EETASANDVAKTA---QAAADGEDGRGEVEETIETALRALREQSQAVETVESLAAEVE 366
QY 223 DNPWFPGKTPAIQAQSLVDQDTATA--TQIE-----KDGNAIGDAYFAGQNASGAVENAKS 275
DB 367 R--IDGIT-ALIEDIAEETNMLNALNASTEAARTGSDGD--GFVAVADEVKDLAEETREQ 420
QY 276 NNSISNIDSA-----KAATATAKTAQIAEAKKFPDSP-ILQEAQWVTOAEKDLKNKP 328
DB 421 AADISEIVDAVTEKAEDASTAIGEVD-AEVERKITRAEGLVRDFAIV-----DEVANVH 475
QY 329 A--DGSVDPNPG---TTVGGSKQOQSSI-----GSRVSMLLDD 362
DB 476 AVQEISADQGAQSYTDVGVGVVEEAVSYSEETAESDTVAADNAEQTDATDEVDQMD 535

Query Match 7.6%; Score 171; DB 2; Length 993;
Best Local Similarity 24.8%; Pred. No. 0.095;
Matches 102; Conservative 62; Mismatches 151; Indels 96; Gaps 20;
QY 76 GIAS---SNSSSTSRSDVDSTTA---TAPTPTPTSDDYKTAQATAYDTIFTSTSLA 128
DB 366 GVASTPPSAPSGDTAPNNVNTQNTAPNSNAPVSTTPOSTSGKKDGOSFADITTVQVSTA 425
QY 129 D-----LQALVSLQDAVNIKDT---AATDEETA-----IAAEWET-- 162
DB 426 NENTQNTIDKDVKSNEAALTSLSLSNNLDTQAKAAQKQSQALRNISYGLASDKFSDF 485
QY 163 -KNADAIKVGAQITELAKYASDQAILDSLGLK---TSFDLLQALLOSANN--NKAAE 216
DB 486 RESLDNVKSGLEYT--TQY---NQOFIDTLEIKENNVNDSKEIDKVKAAANNRINSLR 540
QY 217 ILKENMODNPVPGKT-PAIAQSLVDQ---TQATATQ-----TEKQNA-----IGDAY 260
DB 541 LVNQI--SNALKNSSSGTAEATKLLDQLSKLSSLSRFDYKVKDLNLSLVSISQIMDEL 599
QY 261 FAGQNASGAVENAKSNNSISNIDSAKAAIATKTAQIAEAKKFPDPSPIQEAQWVIOAE 320
DB 600 NKGQATALSNVQ--SKLNTIDQVINSQAILKNCKTFIDRLQ-----TVLPSTEQQYISAV 652
QY 321 KDLKNIKPADGSDVPNPGTTVGGSKQOQSSISGSIHVSMLLDDAENETASI-----LMSG 374
DB 653 KNAQANFPKVSQDAKAAANFVRNDLPQ-----LEQRLTNATASVKNKLPFLTLNG 701
QY 375 FROMTHMNTENPDSQAQOEELAAQARAAGKAAAGDDSAALADAKALEAA 425
DB 702 YDQAVGLLNKNQQAOKKALSILA-----DFSQNKLPDVEKDLKA 741
```


Fri Feb 8 08:45:30 2002

us-09-391-606-9.rspt

Page 9

Search completed: February 7, 2002, 21:41:19
Job time: 4648 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:33:50 ; Search time 12230.8 Seconds
(without alignments)
1923.418 Million cell updates/sec

Title: US-09-391-606-12
Perfect score: 1426
Sequence: 1 tggagcagctactgtgtgtgc.....tccttgaccctggaagtgc 1426

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues
Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1098	77.0	1170	1	AF131889	AF131889 Chlamydia
2	1098	77.0	1170	15	AE033787	AE033787 Chlamydia p
3	1098	77.0	1830	1	CHTMOMPP	M69230 Chlamydia p
4	1098	77.0	20057	1	AE001652	AE001652 Chlamydia
5	1098	77.0	25150	1	AE002167	AE002167 Chlamydia
6	1098	77.0	300550	1	AP002547	AP002547 Chlamydia
7	1098	77.0	1170	1	CHTMOMPB	M64064 Chlamydia p
8	1066	74.8	1602	1	CPMOMPX	X72023 Chlamydia p
9	997.2	69.9	1170	1	CHTMOMPEQ	L04982 Chlamydia p
10	967	67.8	999	1	CHTMOMPAEI	M73038 Chlamydia p
11	668.4	46.6	670	1	AF131230	AF131230 Chlamydia
12	665.2	46.6	670	1	AF131229	AF131229 Chlamydia
13	606.6	42.5	1425	1	AF269259	AF269259 Chlamydia
14	600.2	42.1	1170	1	AF272945	AF272945 Chlamydia
15	600.2	42.1	1261	1	CHTMOMPAAD	M73036 Chlamydia p
16	600.2	42.1	1337	1	CPOMP	X51859 Chlamydia p
17	600.2	42.1	1338	1	CHTMOMPAANA	L39020 Chlamydia p
18	595.8	41.8	1435	1	AF269256	AF269256 Chlamydia
19	595.8	41.8	1436	1	AF269267	AF269267 Chlamydia
20	592.6	41.6	1430	1	AF269266	AF269266 Chlamydia
21	591.8	41.5	1392	1	AF269279	AF269279 Chlamydia
22	585.8	41.1	1176	1	CHTMOMPAHA	L25436 Chlamydia p
23	585.8	41.1	1412	1	AF269260	AF269260 Chlamydia
24	585.8	41.1	1421	1	AF269261	AF269261 Chlamydia
25	581.2	40.8	1482	1	AF269282	AF269282 Chlamydia
26	574.2	40.3	1387	1	AF269280	AF269280 Chlamydia
27	573.4	40.2	1200	1	AF269257	AF269257 Chlamydia
28	573.4	40.2	1200	1	AF269258	AF269258 Chlamydia
29	573.4	40.2	1227	1	CHTMOMPAH	AF269258 Chlamydia
30	571.8	40.1	1343	1	CPMPNOMP	M73037 Chlamydia
31	555.6	39.0	1434	1	AF269269	X61096 Chlamydia
32	552.2	38.7	1107	1	CPDNAX	AF269269 Chlamydia
33	543	38.1	1024	1	CHTMOMPAK	Z18756 Chlamydia
34	536	37.6	1058	1	CPAJ5617	M73040 Chlamydia
35	534.4	37.5	1058	1	CPAJ5613	AJ005617 Chlamydia
36	534.4	37.5	1058	1	CPAJ5614	AJ005613 Chlamydia
37	532.8	37.4	1058	1	CPAJ5615	AJ005614 Chlamydia
38	532.8	37.4	1058	1	CPAJ5618	AJ005615 Chlamydia
39	532.8	37.4	1058	1	CPJ004875	AJ005618 Chlamydia
40	529.6	37.1	1445	1	AF269265	AJ004875 Chlamydia
41	529.6	37.1	1610	1	CHTMOMPXX	AF269265 Chlamydia
42	528	37.0	1058	1	CPJ004874	L04980 Chlamydia p
43	528	37.0	1393	1	AF269262	AJ004874 Chlamydia
44	528	37.0	1432	1	AF269264	AF269262 Chlamydia
45	528	37.0	1438	1	AF269263	AF269264 Chlamydia

ALIGNMENTS

RESULT 1

LOCUS

AF131889

DEFINITION

Chlamydia pneumoniae major outer membrane protein (MOMP) gene, complete cds.

ACCESSION

AF131889

VERSION

AF131889.1 GI:4545320

KEYWORDS

Chlamydia pneumoniae

SOURCE

Chlamydia pneumoniae

ORGANISM

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia

REFERENCE

1 (bases 1 to 1170)

AUTHORS

Sriram,S., Mitchell,W.M. and Stratton,C.W.

TITLE

Multiple sclerosis associated with Chlamydia pneumoniae infection of the CNS

JOURNAL

Neurology 50 (2), 571-572 (1998)

MEDLINE

98145402

REFERENCE

2 (bases 1 to 1170)

AUTHORS

Mitchell,W.M., Tharp,A.C., Stratton,C.W. and Sriram,S.

TITLE

Direct Submission

/db_xref="GI:144541"
/translation="MKRLKALLSAAAFAGSVGSLQALPVGNPSPDLSLLIDGTWEGA
AGDPDCATWCDATSLRAGFYGVDFRILKVDAPKTFSGAKPTGSAANVTAVD
RPNPYNKHLHDAEFTNAGFIALNIWDVDFCTILGASNGYIRGNSTAFNLVGLFGV
KGTIVNANLNPVLSNGVVELYDTSFWSVGARGALWECGATILGAFOYQASPKP
VEELNVCNVSQFVSNPKPKYKGVAFPLPDAGVATGTFKSATINHEWQVQASLSY
RLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFS
DFMOIVSQINKFKSRKACGVTVGATLVDAKWSLTAEARLINERAHVSGQFRF"
386..1483
mat_peptide /note="putative"
stem_loop 517 a 383 c 355 g 575 t
BASE COUNT 1502..1528
ORIGIN

Query Match 77.0%; Score 1098; DB 1; Length 1830;
Best Local Similarity 100.0%; Pred. No. 4.2e-294;
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 129 ttgcctgtagggaacccctctgtatccaaagcttatttaattgatgtgtacaaatgggaaggt 188
DB 386 TTGCCTGTAGGGAACCCCTTCTGATCCAAAGCTTATTAAATGTGTTGATGATGGAAGGT 445
QY 189 gctgcagagatcccttgatcgactctgctactgtgctgagcgtattagcttgcgt 248
DB 446 GCTGCAGGAGATCCTTTCGCATCCTTTCGCTACTTGTGCGAGCCTATTAGCTTACGTGCT 505
QY 249 ggattttacggagactattttccacgcgtatcttaaaagttagatgcacctaataacatt 308
DB 506 GGATTTTACGGAGACTATGTTTTCGACCGTATCTTAAAAAGTAGATGCACCTAAAAACATTT 565
QY 309 tctatgggagcaagcctactggatcgctgctgcaactactactacgcttagataga 368
DB 566 TCTATGGGAGCCAAAGCCTACTGGATCCGCTGCTGCAAACTATACACTCCCGCTAGATAGA 625
QY 369 cctaaccgctcacaataagcatttacacgatgagagtggttcaactaatcagcgttc 428
DB 626 CCTAACCGGCTTACAAATGAAGCATTTACACGATGACAGTGCTTCACTAATGACGCTTC 685
QY 429 attgccttaaacatttgggagcgttctgtatgtttctgtacttttagagccttcaatgt 488
DB 686 ATTGCCTTAAACATTTGGGATCGCTTGTATGTTTCTGTACTTTAGGAGCTTCTAATGT 745
QY 489 tacattagagaaactctacagcgttcaactctgttgggtttattcggagttaaagtaact 548
DB 746 TACATTAGAGAACTCTACAGGCTTCACTCCTGTTGTTTATTCGAGCTTAAAGTACT 805
QY 549 actgtaaatgcaaatgaactaccacacgcttcttcttaagtaacgaggttctgaactttac 608
DB 806 ACTGTAATGCAATGAACCTACCAACCGTTCCTTTAAGTAACGAGCTTCTTGAACCTTAC 865
QY 609 acagacacctcttcttggagcgttagcgcctcgtggagccttatgggaatcggtgt 668
DB 866 ACAGACACCTCTTCTCTTGAGCGTAGGCGCTCTGAGCGCTTATGGGAATCGGCTGT 925
QY 669 gcaacttggagctgaattcccaatgacacagtcacacaccccaacacacacacacacac 728
DB 926 GCAACTTTGGAGCTGAATCCCAATATGACAGCTGCAACCTCAACCTTAAAGTGAAGAACTAAT 985
QY 729 gtgactgttaacgtatgcgaattctctgttaacacacacacacacacacacacacacac 788
DB 986 GTGATCTGTACGTATGCAATTTCTCTGTAAACAAACCCAGCGCTTAAAGCGCTGTCT 1045
QY 789 ttcccttggcaacagacgctggtcgttagcagcgtactggaacaaagctctgcgacacac 848
DB 1046 TTCCCTTGGCAACAGACGCTGGGTAGCAACAGCTACTGGAACAAAGTCTGCGACCATC 1105
QY 849 aattcatgatgaggaagtaggagcctctctatttcaagacataaactcttagtgcca 908
DB 1106 AATTATCATGAATGGCAAGTAGGAGCCTCTCTATCTTACAGACTAACTCTTTAGTGCCA 1165
QY 909 tacattgagtagaattgctcagacacattttagctgataacatccgactgctcag 968

DB 1166 TACATTGGAGTACAAATGGTCTCGAGCAACATTTTGTATGCTGATAACATCCGATTCAG 1225
QY 969 ccaaaactacactacagctgttttaacttaacttaactgcagtggaaccccttcttactaggaat 1028
DB 1226 CCAAACTACCTACAGCTGTTTAACTTAACCTTAACCTGCAATGGAACCTTCTTTTACTAGGAAT 1285
QY 1029 gccacagcattgtctactactgattgcttctcagacttcacaaattgtttctcgtcag 1088
DB 1286 GCCACAGCTTGTCTACTACTGATTCTGTTCTTCAGACTTCATGCNAATGTTTCTCTGTCAG 1345
QY 1089 atcaacaagttaaatctagaaaagcttggaggttacttagagcgtacttagttagt 1148
DB 1346 ATCAACAGTTAATCTAGAAAAGCTTGTGGAGTTACTTGTAGGAGCTACTTTAGTTGAT 1405
QY 1149 gctgataaattggtcacttactactgcagaagcctgttttaacttaacagagagcgtgcacgta 1208
DB 1406 GCTGATAAATGGTCACTTACTTGCAGAAAGCTCGTTTAAATTAACGAGAGAGCTGCTCACGTA 1465
QY 1209 tctgctcaggttcagattc 1226
DB 1466 TCTGTCAGTTCAGATTC 1483
RESULT 4
AE001652 20057 bp DNA BCT 01-DEC-2000
LOCUS Chlamydia pneumoniae section 68 of 103 of the complete genome.
DEFINITION AE001652 AE001363
ACCESSION AE001652 AE001363
VERSION AE001652.1 GI:4376997
KEYWORDS
SOURCE Chlamydia pneumoniae CWL029.
ORGANISM Chlamydia pneumoniae CWL029
REFERENCE 1 (bases 1 to 20057)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 2 (bases 1 to 20057)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L., Grimwood, J., Davis, R. W. and Stephens, R. S.
TITLE Direct Submission
JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
FEATURES
source 1..20057
/organism="Chlamydia pneumoniae CWL029"
/strain="CWL029"
/db_xref="taxon:115713"
339..1508
/gene="ompA"
/note="Cpn0695"
339..1508
/gene="ompA"
/codon_start=1
/transl_table=11
/product="Major Outer Membrane Protein"
/protein_id="AAD18834.1"
/db_xref="GI:4376998"
/translation="MKRLKALLSAAAFAGSVGSLQALPVGNPSPDLSLLIDGTWEGA
AGDPDCATWCDATSLRAGFYGVDFRILKVDAPKTFSGAKPTGSAANVTAVD
RPNPYNKHLHDAEFTNAGFIALNIWDVDFCTILGASNGYIRGNSTAFNLVGLFGV
KGTIVNANLNPVLSNGVVELYDTSFWSVGARGALWECGATILGAFOYQASPKP
VEELNVCNVSQFVSNPKPKYKGVAFPLPDAGVATGTFKSATINHEWQVQASLSY
RLNSLVPYIGVQWSRATFDADNIRIAQPKLPTAVLNLTAWNPSSLGNATLSTDSFS
DFMOIVSQINKFKSRKACGVTVGATLVDAKWSLTAEARLINERAHVSGQFRF"
complement(173..1803)
/product="rRNA-Gly"
1892..2725
/gene="rs2"
/note="Cpn0696"

```

CDS
1892..2725
/gene="rs2"
/codon_start=1
/transl_table=11
/product="s2 Ribosomal Protein"
/protein_id="AAD18835.1"
/db_xref="GI:4376999"
/translation="MESQCKLTIKDLSAGAHGHOTRRNPKMKLYIFEEKNGLYI
INAKTQALRALPHIRKVIQDNKTVLFVGTKKQAKCVIREAAIEAGEFFIAERWLG
GLMTNTTIRSIKTDKIEDLSRNOAYLTKKEALLAKRHOKLLRNLEGIRYMKKA
PGLLVVDPSPSEKIAVAEAKKLGIPVALVNTCPDTPIDHVICNDDSLKSIRLIIN
VIKENIEAKHKLGIETVSPKSLVPPDLSAFESSQDDDESDEENREEDLLAKKFDCEA
N"
2725..3573
/gene="tsf"
/translation="CpN0697"
2725..3573
/gene="tsf"
/codon_start=1
/transl_table=11
/product="Elongation Factor TS"
/protein_id="AAD18836.1"
/db_xref="GI:4377000"
/translation="MSDFSMTLTKLROQTGVLTKCKEALACGNGLEAAVYLRKL
GLSAGKKEHRTKEGIIAAKTDANGTALIEVNETDFVANNVAFREVFVNLLDILK
YKVDTEALSAQSPDLSVDELRAVTMQTVGENIRISVAYFPKATNSTVGIYSH
NGKTVATMLSGSTADSLAKDIAMHVVAQPOFLSKESVPAEAIKAEKEVIAEQIQ
GKQPEVIEKIVTGKLTFFQEAQLLEQPFKINADLSIQSLIDDFSKTSGSSVAIEQFI
LWKICA"
3581..4327
/gene="pyrH"
/translation="CpN0698"
3581..4327
/gene="pyrH"
/codon_start=1
/transl_table=11
/product="UMP Kinase"
/protein_id="AAD18837.1"
/db_xref="GI:4377001"
/translation="NAKQTRRVLFKISGEALSQDSNRIDEMRLSRLVSELRAVRNND
IEIALVIGGNLRLGAQKELQINRVADQMGLATLNGMAVADAKAEDIPCLLT
STMSCPQADLVTPOKSIEALDOGKILICTCAGSPYLTTDTGAALRACELNVDVLIK
ATMHVGVYDKDPRLPDPAKYDFVSKDFLTSNQLGVMDASALSICMDSHIPRVSF
LOHSLKALFDFTICTGLVSEDVNVHVCSPRH"
4305..4847
/gene="trf"
/translation="CpN0699"
4305..4847
/gene="trf"
/codon_start=1
/transl_table=11
/product="Ribosome Releasing Factor"
/protein_id="AAD18838.1"
/db_xref="GI:4377002"
/translation="MSVLQDTEKKMAALDFHKEVKFSFRCKAHPALVETVVVDYG
TWRUSDIASISVADRLQVISPYNQNASAIKAGIIAANLNIQPEVSGSIIRKVE
FPADYQEMIQLRKEEAKINVRIRREANDKLKDSALTEDVYVGNKKEIKQLD
FKQKDELTKKEAEIASI"
4945..5019
/product="trNA-Glu"
complement(4945..5019)
/translation="CpN0700"
5220..5735
/product="trNA-Lys"
complement(5045..5117)
/translation="CpN0700"
5220..5735
/gene="CpN0700"
/codon_start=1
/transl_table=11
/product="Cp676 hypothetical protein"
/protein_id="AAD18839.1"
/db_xref="GI:4377003"
/translation="MVHSPTHOCYHCQOPATICYTEIDKDKVIRSYVCATCPSPSHY
NNEHLSKSGVGLTLEGCNCKTVVHSHKQDDQLLGCHQCITYNFKNQITSLKSERV
SSSFTMERGSGSLHIGRAPGEASNTNPLLLKIALNEALQDTLEREDYEQAAVIRQIN
HLTKNPDPS"
5722..6798
/gene="kargC"
/translation="CpN0701"
5722..6798
/gene="kargC"
/codon_start=1
/transl_table=11
/product="Arginine Kinase"
/protein_id="AAD18840.1"
/db_xref="GI:4377004"
/translation="MTPLNDELTLVLRKESQANKVMPVTFTSLARNLSVSKFLPCL
SKEQLEITQITTSFHNIEGFEFIVLPKDTPLMKFELFLEHFLPYDLVGNPEGE
ALVVSRSQDGLAAINFQDHLVIGDFQGNVEKTDQLQVLDSDYLSHSLSAFSEFG
FLTTPNKGCTGLKSOCLHPIPALYKFEFTNLIDEEVEITTSLLGVGTGPPGNIVV
LSNCSGLTRELSSLRITASKLSVAEVAARKRLSEENSQDKNLLRLSLGLLTHS
CQELKETLDALSWIQLDIGLILKIVTENHPLNPLFWQIRAHALQKQAEQSDQLQ
KDTISHLRASVLKELTKGLSPESF"
complement(6831..6903)
/product="trNA-Thr"
complement(7049..9808)
/gene="yscC"
/translation="CpN0702"
complement(7049..9808)
/gene="yscC"
/codon_start=1
/transl_table=11
/product="probable Yop proteins translocation protein
C/general secretion pathway protein"
/protein_id="AAD18841.1"
/db_xref="GI:4377005"
/translation="MKTVLNIGRKILQIOLIKKKKIKIGLSGLFFLDLVLLGVSSQRP
TETSANVKNLRDEKLAACPKNASASLSAKSKTKTTPGSTPSKVSFKFADQDKTF
QKTSQSAFPKPTTLKELEERKPPERTADVKRSRPRELPTQVEVPVPAKSEQL
DSIQWEEKQNTARRAVNAINLSIKKLEEQSTVTEKDVQKQATPHASKNNVASP
STMQVEIKAAATTAAPQDKSEERKELTKRELTEDLDNGVTYVFEDISILELL
QVSKISGTFNFDSDLNQFNVTIVSHDPTSYDDLSITLLQVKKHDLKVVQGNVL
IYRNPHLSKLSVTVDSSLTCEAVVTVRFLYSVPSAAVNIQPLSHDAIVSA
SEATHFIISDIAGNVKVDLLAALDCPGTSVDMTEVEVKYANFAALVSYQDVLGT
LAEDDAVQIQTPGNKIFVYSSPLANKAQLKSLDVPMAHTLDDPASLALGG
TGTSFSLRFEMFKLYQNGEVIANALQDIGNLYVTTAMDEDINTLNSIQWLEVN
NSVIATIQGNQNDVRVIGLNGLDLPPKQVYIEVLIDTSLSKSWDFQGVQVVALGDEQS
KVAYASGLNNNTGIATPTKATVPPTPNPSPITPQGLTQFSDMLNSSAFGLGII
GNVLSHGKSELTGLGLSALDQDGTIVLNPRIQAQDQQAQSFVGTQVPTQNT
IIQETGTVQIDIEDIGVNLVTVSTVAPNNVTLQIEQITSELHSGSLTPVTDKT
YAATRLQIPDGCFLVMSGHIRDKTTKVYSGVPLNLSIPLIRGLFSRTIDQOKRNMIM
ETKPKVISSEFEGTATNKEGYRNYWEADEGSMQVAPRAHPECQGPSPSLQAESDFKII
EIEAQ"
complement(9805..11313)
/gene="CpN0703"
complement(9805..11313)
/gene="CpN0703"
/codon_start=1
/transl_table=11
/product="S/T Protein Kinase"
/protein_id="AAD18842.1"
/db_xref="GI:4377006"
/translation="MDCRGITLPPEQVIGVHVKKILSKLRSRVVHGLHPETRHST
VIKVPSPSPSTSRVYNFLKEAQLSHOITHRNIVKFRYCKWQDCLYIAMEYIEGIS
LREYILAQFISLPQADIIIDIAQALEHLHSNLIHDKIPENILITPQKTKLIDFG
LADWDEIQRAPHSVIGTPYNSPEQSGESHPASDIYALGLLAYELTLGLHSLGRV
FLSLVPERISKILAKALQSPNNRYSSTREFIQDIIHHYRMSQDMQEDIRIKDHTVALY
BOLOTQRFELAPETLPFPDFISGVLYHQGYPLYPHAYDTLLEGVYVNLWGLYSPISNA
TIALSVKSLVCOODLQRLPDLDRVCEINEELRMKIPIDEMGICISILCLEISKENKELS
WIACGTVFWIKRQGVQDPEFSGGLGKTLQIURETKVAMEIGDEAVVCTLELE
SVASLKTLSLAEQLDRRKAIFCPIESIHGGIQSRQSGNSPSTLISLKRIN"
complement(11329..12444)
/gene="flin"
/translation="CpN0704"
complement(11329..12444)
/gene="flin"
/codon_start=1

```

```
/transl_table=11
/product="Flagellar Motor Switch Domain/YscQ family"
/protein_id="AADI8843.1"
/db_xref="GI:4377007"
/translation="MAAADSSAHLKSRNLFSSLGKTEEOVAAPFEKPELCOHKIR
EKFRLEDVQVSIKFRSGITAVEATKEFGVHLLIQPVVQVPEVENLLFLTSEEDLQEL
```

Query Match
Best Local Similarity 100.0%; Pred. No. 5.5e-294;
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 129 ttgctgtagggaaccccttgcacccaagcttattatgtatgatacaataatggaaggt 188
DB 408 TTGCTGTAGGGAACCCCTTCTGATCCCAAGCTATTATTAATGATGGTACAAATGGAAGGT 467
QY 189 gctgcagagatccctgcgactctgcctgctgctgctgctgctgctgctgctgctgct 248
DB 468 GCTGCAGGAGATCCTTGGGATCCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 527
QY 249 ggatttaaggagactatgttttgcacccgtatctttaaagtagatgacccctaaacattt 308
DB 528 GGATTTTACGGAGACTATGTTTTCACCGGTATCTTAAAGTAGATGCACTTAAACACTTT 587
QY 309 tctatggaagcaagcctactgagctgctgctgctgctgctgctgctgctgctgctgctgct 368
DB 588 TCTATGGGAGCCCAAGCCTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 647
QY 369 cctaaccggcctacaataagcattacacagatgacagagtggtgctgctgctgctgctgctgct 428
DB 648 CCTAACCCGGCCTACAAATGAAGCATTTACAGATGACAGATGCTGCTGCTGCTGCTGCTGCT 707
QY 429 attgcttaaacatttgatgctgctgctgctgctgctgctgctgctgctgctgctgctgct 488
DB 708 ATTGCCTTAAACATTTGGGATGCTTGTGATGTTTCTGCTGCTGCTGCTGCTGCTGCTGCT 767
QY 489 tacattagaggaactctacagctgctgctgctgctgctgctgctgctgctgctgctgctgct 548
DB 768 TACATTAGAGGAACCTCTACAGCTTCAATCTCGTTGTTTATTCGGAGTTAAAGGTACT 827
QY 549 actgtaaatgcaatgcaactacaacgcttctttaaagtaacgagtggtgtaactttac 608
DB 828 ACTGTAAATGCAATGAATACCAACGTTCTTTAAGTAACGGAGTGTGTAACCTTTAC 887
QY 609 acagacacctcttctcttggagcgttagcgtcgtgagccttattggaatgcggtgtg 668
DB 888 ACAGACACCTCTTCTCTTGGAGCTGAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 947
QY 669 gcaacttgggagctgaattccaatgacagctgctgctgctgctgctgctgctgctgctgct 728
DB 948 GCAACTTTGGGAGCTGAATTCCAATATGCACATGTCACACCTTAAAGTTGAAGAACTAAT 1007
QY 729 gtgactgttaacgtatcgcaattctctgtaacaaaccccaaggctataaaggcgttgt 788
DB 1008 GTGATCTGTACGATATGCGAATCTCTGTAAACAAACCCAGGCTATAAAGCGTGTCT 1067
QY 789 ttcccttggcacaagacgctgctgctgctgctgctgctgctgctgctgctgctgctgctgct 848
DB 1068 TTCCCTTGGCAACAGACGCTGGCGTAGCAACAGCTACTGGAACAAAGTCTGCGACCATC 1127
QY 849 aattatcatgaatggcaagtagagcctctcttcttctacagactaaactcttagtgcca 908
DB 1128 AATTATCATGAATGGCAAGTAGGAGCTCTCTATCTTACAGACTAAACCTTTTAGTGCCA 1187
QY 909 tacattgagtaaatggtctgacgaacttttgcgtgctgatacaatccgcattgctcag 968
DB 1188 TACATTGGAGTACAAATGCTTCGAGCAACTTTTGTGATGCTGATAACATCCGCTGCTCAG 1247
QY 969 ccaaaactacacagctgttttaaaccttaactgaatggaacccctcttcttactagaat 1028
DB 1248 CCAAACTACCTACAGCTGTTTTAACTTAACTGATGCAACCTCTTTTACTAGGAAT 1307
QY 1029 gcaacacattgctactactgattgctgctgctgctgctgctgctgctgctgctgctgctgct 1088
```

```
DB 1308 GCCACAGCATTTCTACTACTGTTCTCTCAGACTTCTCAGCAATTTGTTCTCTGTCAG 1367
QY 1089 atcaacaagttaaatatagaaaagcttctgagagtagctgtagagctactttagtgat 1148
DB 1368 ATCAACAAGTTTAAATCTAGAAAGCTTGTGGAGTTACTCTAGAGCTACTTTAGTTGAT 1427
QY 1149 gctgataaagtgcactactcagagctcgttttaataacagagagctgctcacgta 1208
DB 1428 GCTGATAAATGGTACTTCTGAGAGCTCGTTTAAATTAACGAGAGCTGCTCAGCTA 1487
QY 1209 tctggtcagttcagattc 1226
DB 1488 TCTGTCAGTTTCAGATTC 1505
RESULT 5
AE002167/c 25150 bp DNA BCT 06-APR-2001
LOCUS Chlamydomophila pneumoniae AR39, section 4 of 94 of the complete
DEFINITION genome.
ACCESSION AE002167 AE002168 AE002161
VERSION AE002167.2 GI:8163349
KEYWORDS
SOURCE Chlamydomophila pneumoniae AR39.
ORGANISM Chlamydomophila pneumoniae AR39.
REFERENCE 1 (bases 1 to 25150)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
Nucleic Acids Res. 28 (6), 1397-1406 (2000)
JOURNAL
MEDLINE 20150255
PUBMED 10684935
REFERENCE 2 (bases 1 to 25150)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., Deboy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Direct Submission
Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA
On Jun 1, 2000 this sequence version replaced gi:7188982
gi:7188971.
FEATURES
source Location/Qualifiers
1..25150
/organism="Chlamydomophila pneumoniae AR39"
/strain="AR39"
/db_xref="taxon:115711"
/note="synonym: Chlamydia pneumoniae AR39"
306..1325
/gene="CP0032"
306..1325
/gene="CP0032"
/note="similar to GB:M25323 SP:P13580 GB:M30785 GB:X17434
PID:146330; identified by sequence similarity; putative"
/codon_start=1
/transl_table=11
/product="glutamyl-tRNA reductase, putative"
/protein_id="AAF37927.1"
/db_xref="GI:7188972"
translation="MVLGVGVISYREALKERERAIOYLOSFEKLNFLAORFLKGGGA
FIPLLTCHRAELYYSEPEIAQAALLSELTSQGRPYRHRGLSCFTHLFVQTSIDS
LIFGTEIQGOVKRAYLKGSKERELPDLHFLFKALKKEKYEYSRIGFPPHQVTIES
VYQELLSDYSIYTNFLVGYSDINRKVAALYQHGVRITFCRSQOVTPAYRTLSR
ETLSFRPYDVIFFGSSSESASFSDLSCEASIPKRIYVDFNPVPTFLMKETPTGFV
YLDIDFISECVOKRLQCTKEGVNKAULLLTCAAKKQWEIIEKKSHITQRIQISSRIP
SVLSY"
1692..2084
gene
```



```

CDS
/ gene="CP0033"
1692..2084
/ gene="CP0033"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37928.1"
/ db_xref="GI:7188973"
/ translation="MLEKLIKNFATYMGITITLLELDADGAVLPISEVVKVRAQONAD
NEIVLSASIGALPSADPAKLYLQMIGNLFRGRETGGSGALDSEGNVVMVRRFSGDT
TYDDFVRHVESFMNFSETWLSDLGLGQ"
2098..4635
/ gene="CP0034"
2098..4635
/ gene="CP0034"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37929.1"
/ db_xref="GI:7188974"
/ translation="MAVRLIVDEGPLSGVIFVLEDIGSISGRSSANDIPIEDPKLG
ASQAIINTKDGSYITNLDSDTIPIVWNGVAIQETTLKNEDTILLGNOYKFLSDGSD
PQDLVYDFDIPEENFSNDGLSDSNEOGKLEPROTSETNHSKPKEKLFKDGSSD
PITSGDQELADFLASAKAEKNOPRAKYAKGLKESNESLNPKQNAKDSPEERT
NKQNAIMEDNGASPRQDPQPKSAEPLSKNTARDETPUKENPVEEKANKATDPSPE
KKDQPEGSKGSKIATPLDSQKESDEKAEAEFYQEEENETEDNKEDSAAADA
NDDTASHTADENKETPKVNEKSAVLSFPHVDLFRDQTIFFAEIDDIAGKNISV
DLTOPSRELLKLAGANIGAEFHLDSGTYILGTDPTCDIVFNDLSVSHQAKITVG
NDGGTLEIDLSKNGVIVEGRKIDKSTLSSNOVVALGCTFLLLIDHAPADTIVASL
SPDDYSLFGRODAEALERQAEKQKRAIPLAGSFILTLFVGGILLEGITAS
LEHFKVYPLNIDYQELDAQVINGFPVRYTPFKNTNSQLFLIGHVKNSTDKSELLYK
VDALSVKSDNVNIDDEAVQWQEMNILLKRPFGISMHSPEKFIITGYVKTTEQ
AACLDYDINIHFNLNENKVVVETOMLKAAGHLQGGSPHAFVAVNGEVLITGY
VNDDAEFRVAVQELSGIPGVRLVKNFAVLLPAEGLIDLNLPNRYRVTGYSRG
EISINVVNGRILTRGDVIDGMTVTSIQPNAIFLEKEGLKYKIDYNK"
4682..4930
/ gene="CP0035"
4682..4930
/ gene="CP0035"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37930.1"
/ db_xref="GI:7188975"
/ translation="MFNMNTAKEKNQSPLLDLQDMQDHDRAQELKASVDQKVHKL
HALLREGSDKESFGQQSLLAGYVALQKVLGRINKMI"
4952..5206
/ gene="CP0036"
4952..5206
/ gene="CP0036"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37931.1"
/ db_xref="GI:7188976"
/ translation="MATNKSCTAFDNKMLDGCVTYVRGVOQYLTELETSTQGTVDLG
TMFNLFQRMQILSQYMESVSNLLTAVNTEMITMARAVKS"
5225..5674
/ gene="CP0037"
5225..5674
/ gene="CP0037"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11

CDS
/ product="conserved hypothetical protein"
/ protein_id="AAF37932.1"
/ db_xref="GI:7188977"
/ translation="MADLEVFQADFALLFEAGLIAIKOGDEDSARKLQPSUHLINPNH
YCHDGLALISLHKMDLFDFAERLSALIKGNEKDNMSTRKFLSLTHMLIVLHOGSSFEV
RRSLESLCKRFADQVIANCKIESTRALAQSVLDNHDITLVAKSAGPLG"
5708..6382
/ gene="CP0038"
5708..6382
/ gene="CP0038"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37933.1"
/ db_xref="GI:7188978"
/ translation="MIDPEVECPNLDGDAEQSITQNSGTPLASELKKDIPFALGSY
AAPKDTTLVQCFKPNMAMQDQNSLIDPELOALESELEQEOINNLKGRLWDFRST
FEDSOTTAQFADEHFOAVGVIIDILNEDLNTIAEHTQODARKDEKESVTRKIIDWV
SSGSEVLNALLYPSDRDGNRESLANFLKVOYAVQRATQRAELFASIVGTSVSSVKTI
MITQLG"
6384..7712
/ gene="CP0039"
6384..7712
/ gene="CP0039"
/ note="similar to SP:P21211 GB:U02499 SP:P21212 SP:P40290
PID:437202; identified by sequence similarity; putative"
/ codon_start=1
/ transl_table=11
/ product="type III secretion cytoplasmic ATPase Sctn"
/ protein_id="AAF37934.1"
/ db_xref="GI:7188979"
/ translation="MDQLITDFDILMSQLGDVNLTVVGRITVEVGMILKAVPNVRV
GEVLCVLRNGMEPLTVVVGTSQFAFLSGELSGVSPSSSEVITPLHLIRAGNGL
LGRVLNIGEPIDIVETKGLQNVQTFPIFRAPDPLHRAKLQILSTGVRCIDGMLT
VARQIRIGIFAGAGYKSSLLGMIAENAEADVNVIALIGREGRVRFIEGDLGEEG
MKRSVIVYSTDSQSLRNAAYTAETAEYFROGKTVLWLMDSVTRFARALREVGL
AAGPPPARAGVTPSVFSTPLRLPERSGASDKGTITAFYTVLVAGDMNPEVADEKSI
LDGHVLSNALAQAYHYPADIVLASISRLLLTAIYPEEQRRIIGKAREVLAKYKANEML
IRIGEYRGRSDREIDFALDHIDKLNRFUKDIHEKNTYEAAQQLRAIFR"
7730..8236
/ gene="CP0040"
7730..8236
/ gene="CP0040"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37935.1"
/ db_xref="GI:7188980"
/ translation="MAKYPLEPVLAIKKDRVDRAEKVVKEXRRLLIEQEKLEKEAE
RDKVKNHYMQIKQLRLDEGTTSDAVLQIKSYKVVAVQLVSEEEKVKNQKEVYLLA
ASKELEKAEVNLAKRRKEEKTRLHKEEMKEALKEAEKEEKEQDMLFQURQK
KKRESGGS"
8240..9082
/ gene="CP0041"
8240..9082
/ gene="CP0041"
/ note="conserved hypothetical protein; identified by
Glimmer2; putative"
/ codon_start=1
/ transl_table=11
/ product="conserved hypothetical protein"
/ protein_id="AAF37936.1"
/ db_xref="GI:7188981"
/ translation="MELKKTAEVLSYSAKTDNHTVYQNSPEPRDSRVKVSLEGKQTR
QEKTTSSKGNTRTESRFADEKRVDDIEAEVSGKEEQSEQFCFLAENAFAGMLID
IAAAGSAAEVVEVAPIAVSSDITQWENIILSTVESWVISEINGQLVELVDASSV
PEAFVGANLTLVQSGDLSVKFSFVDATQMAEADLTNNPSQLSSLSVALKGHQLT

Query Match 77.0%; Score 1098; DB 1; Length 25150;
```

Best local similarity 100.0%; Pred. No. 5, 6e-294;
Matches 1098; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 129 ttgctgttaggaacccctctgataccagcttataatgtaggtacaataatggaaggt 188
|||||
Db 21128 TTGCTGTAGGAACCCCTCTGATCCAGCTTATTAATGTAGTACAAATATGGAAGGT 21069
QY 189 gctgagagatccttgatcccttgctgactacttggtgacgctattagcttcgct 248
|||||
Db 21068 GCTGAGAGATCCCTTGGATCCCTTGGCTACTTGGTGGACGCTATTAGCTTACGTGCT 21009
QY 249 gattttcagagactatgtttgacccgttatcttaaaagtagatgcacctaataacattt 308
|||||
Db 21008 GATTTTACGAGACTATGTTTTCACCGTATCTTAAAGTAGATGCACCTAAACATTT 20949
QY 309 tctatggagccaagcctactggatcccgctgctgcaactactactactgacgtagataga 368
Db 20948 TCTATGGAGCCCAAGCTACTGGATCCGCTGCTGCAAACTACTACTGCGGTAGATAGA 20889
QY 369 cctaaccgacctacaataagcatttacagatgagagtggttcaactaaatgcaggcttc 428
|||||
Db 20888 CCTAACCCGGCTACAATAAGCATTTACAGATGCGAGTGGTTCACCTAATGCAGGCTTC 20829
QY 429 attgcttaaacatttgggagctgcttggatgtttctgtacttttaggagcttctaattgt 488
|||||
Db 20828 ATTGCTTTAAACATTTGGATCGCTTGTATGTTTCTGTTTACTTTAGGAGCTTCTAATGCT 20769
QY 489 tacattaggaatacctacagcttcaactctgctgtgtttatttcgaggttaaaagtact 548
Db 20768 TACATTAGAGAACTCTACAGCTTCAATCTCGTGTGTTTATTCCGAGTTAAAGGTACT 20709
QY 549 actgtaaatgcaatgaaactacacaaacgctttcttaagtaacgaggtgttgaactttac 608
Db 20708 ACTGTAATGCAATGAACACTACCAACGTTCTTTTAAGTAACGAGTGTGTTGAACCTTAC 20649
QY 609 acagacacctcttctcttctgagcgttagcgcctgtgagaccccttatgggaatgcggttgt 668
Db 20648 ACAGACACCTCTTCTTCTTGGAGCGTAGCGCTGCTGAGGCTTATGGAATGCGGTTGT 20589
QY 669 gcaactttggagctgaattccaattacacagctcaaacctcaaaagttgaagaacttaat 728
Db 20588 GCAACTTTGGAGCTGAATTCGAATGACACAGCTCAACACCTTAAGTTGAAGACTTAAT 20529
QY 729 gtatctgtaacgtatcgcaattctctgttaacacacccaaagggctataaaggcgttgt 788
Db 20528 GTGATCTGAACGTATCGCAATTCCTGTAAACAAACCCAGGCTATAAAGCGTGTCT 20469
QY 789 ttccctcttgcaacagacgctggcgttagcaacagctactggaacaaagctctgcgaccatc 848
Db 20468 TTCCCTCTTGCAACAGACGCTGGCGTAGCAACAGCTACTGGAACAAAGCTCGGACCATC 20409
QY 849 aattatcatgaatggcaagtagagcctctctctattctacagactaaactcttagtgcca 908
Db 20408 AATTATCATGAATGGCAAGTAGAGGCTCTCTATCTACAGACTAACTCTTTAGTGCCA 20349
QY 909 tacattggagtaacattggtctgagcaacttttgatgctgataacacccgactgctcag 968
Db 20348 TACATTGGAGTACAATGGTCTCGAGCAACTTTTGTATGCTGATTAACATCGCAATTGCTCAG 20289
QY 969 ccaaaactactacagctgttttaacttaactgaactgaacccctcttacttaggaat 1028
Db 20288 CCAAAACTACTACAGCTGTGTTTAAACTTAACGTGATGAACCCCTCTTTACTAGGAAT 20229
QY 1029 gccacagcattgtctactactgattgcttctcagactctcatgcaaatgtttctcgtcag 1088
Db 20228 GCCACACATGCTACTACTGATTCGTTCTCAGACTTCAATGCAAAATGTTTCTCCTCAG 20169
QY 1089 atcaacaagtttaaatatagaaaagcttggaggttactgtgagagctacttagtgtat 1148
Db 20168 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTTACTGTAGAGCTACTTGTAGTGTAT 20109
QY 1149 gctataaatggctactactcagaaagctcgttttaataacagagagcgtcgcagta 1208
|||||

Db 20108 GCTGATAAATGGTCACCTACTGAGAGAGCTGTTTATTAACGAGAGAGCTGCTCAGTA 20049
QY 1209 tctgggtcagttcagattc 1226
|||||
Db 20048 TCIGGTCAGTTGATTC 20031
RESULT 6
AP002547 300550 bp DNA BCT 08-DEC-2000
LOCUS Chlamydomophila pneumoniae J138 genomic DNA, complete sequence,
DEFINITION section 3/4.
ACCESSION AP002547 AB033786 AB033787 AB033816 AB033817 BA000008
VERSION AP002547.2 GI:10176693
KEYWORDS
SOURCE Chlamydomophila pneumoniae J138 (strain:J138) DNA.
ORGANISM Chlamydomophila pneumoniae J138
REFERENCE 1 (sites) Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K.,
Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.
Comparison of whole genome sequences of Chlamydia pneumoniae J138
from Japan and CWL029 from USA
Nucleic Acids Res. 28 (12), 2311-2314 (2000)
JOURNAL 2030349
MEDLINE Direct Submission
REFERENCE 2 (bases 1 to 300550)
AUTHORS Shirai,M.
TITLE Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.
JOURNAL Mutsunori Shirai, Yamaguchi University School of Medicine,
Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi
755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,
Tel:81-836-22-2227, Fax:81-836-22-2415)
On Sep 15, 2000 this sequence version replaced gi:6172298
gi:6172300 gi:6172396 gi:6172398 gi:8978889
AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).
FEATURES
source
1. 300550
/organism="Chlamydomophila pneumoniae J138"
/strain="J138"
/db_xref="taxon:138677"
/note="synonym:Chlamydia pneumoniae (strain J138)"
complement(21..794)
/gene="dapF"
complement(21..794)
/gene="dapF"
/codon_start=1
/transl_table=11
/product="diaminopimelate epimerase"
/protein_id="BAA98725.1"
/db_xref="GI:8978890"
/translation="MAFYSPSTISKYFIYSGAGNRFLIGETLPEVEDVRFCLQETRV
GFLYKPFSSCADAQLIIFNSDSRPTWCGNLCAIAHLASQKGSIDISVSDGLYS
GYFYSDRLVDMTLADWRASVHRLSRPDLPRKVVCIHTGVPHAVVILPEISTLDL
SILGFLRYHOTFPDGVNPNVFOILGHCOLRVYERGVETAACTGACGTGALASALV
SNSYGWKESIQIHTWGGELMTVSNQRGVYLGVSVTRDL"
complement(763..1338)
/gene="clpP_1"
complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHKLRDIIIEKLEARRVFFSEVFEKSDAIKKILWYL
ELKDPKPIFVINSPPGSDVAGFAVMDQIKMLTSPVTVTGLAASMGVSLCAAP
GRRATPHSRIMIHOPISGIPITGQATDLDIHAREILKTKARIIDVYVEATNQPRDII
EKAIDRWMTANEAKDPGLDGLIFSNL"
complement(1358..2917)
/gene="glaA"
complement(1358..2917)
/gene="glaA"

Db 179466 GCTGAGGAGATCCTTGGGATCCTTGGCTACTTGGTGGAGCGCTATTAGCTTACGGTCT 179525
Qy 249 ggaatttcagagactatgttttgcacgctatcttaaaagttagatgcacctaataacattt 308
Db 179526 GGAATTTACGGAGACTATGTTTTCGACCGTATCTTAAAGTAGATGCACCTAAACATTT 179585
Qy 309 tctatggagccaagcctactgagatccgctgctgcgaactactactgcgtagataga 368
Db 179586 TCTATGGGAGCCAAAGCTACTGGATCCGCTGCTGCAAACTACTACTGCGGTAGATAGA 179645
Qy 369 cctaaccggtctacaataaagcattacacatcacagatggttcactaaatgcaggcttc 428
Db 179646 CCTAACCCGGCTTACAAATAGCATTTACCAATGAGAGTGGTTCACCTAAATGCAGGCTTC 179705
Qy 429 attgcttaaacatttggatcgcttctgatgtttctgactttaggagcttctaaggt 488
Db 179706 ATTGCTTAAACATTTGGATCGCTTTTGATGTTTCTGCTACTTTAGGAGCTTCTAATGTT 179765
Qy 489 tacattagggaaactctacagcttcaactcgttctggtttatttattcggagtttaaggtact 548
Db 179766 TACATTAGAGAACTCTACACGCTCAATCTCGTTGGTATTTCGGAGTTAAAGGTACT 179825
Qy 549 actgtaaatgcaaatgaactaccacacgcttctttaaagtaacgaggtgttgaactttac 608
Db 179826 ACTGTAATGCAATGAAGTAACACTACCAACGTTCTTTAAGTAACGGAGTTGTTGAACCTTAC 179885
Qy 609 acagacacctcttctcttggagcgttaggcgtcgtgagccttatgggaatcggttct 668
Db 179886 ACAGACACCTCTTCTTCTTGGAGCGTAGGCGTCTGAGGCTTATGGGAATGCGGTTGT 179945
Qy 669 gcaactttggagcgtgaattccaataatgcaagtcgaactcaaacctaaagtgaagaactaat 728
Db 179946 GCAACTTTGGAGCTGAAATCCAAATATGCACTGACAGTCCAAACCTTAAAGTTGAAGACTTAA 180005
Qy 729 gtaacttgtaacgtatcgcaattctctgtaacaaacccaaagggctataaagcggttct 788
Db 180006 GTGATCTGAACGATGATCGCAATCTCTGTAACAAACCCAAAGGCTATTAAGGCGTTGCT 180065
Qy 789 ttcccttggcacaagcgtggtcgttagcaacagctactgggaacaaagctcgcgaccatc 848
Db 180066 TTCCCTTGGCAACAGACGCTGCGTAGCAACAGCTACTGGAAACAAAGTCTGCGACCATC 180125
Qy 849 aattatcatgaatggcaagtaggagcctctctattctacagactaaactcttttagtgcca 908
Db 180126 AATTATCATGAATGGCAAGTAGGAGCCTCTCTATCTTACAGACTAAACTCTTTAGTGCCA 180185
Qy 909 tacattggagtacaatggtctcagcaacttttgcgtgatacaactcgcgattgctcag 968
Db 180186 TACATTGGAGTACAATGGTCTCGAGCAACTTTTGTGCTGATTAACATCCGCAATGCTCAG 180245
Qy 969 ccaaaactactacagctgttttaaaacttaactgcatggaaccccttcttactaggaaat 1028
Db 180246 CCAAACTACTACAGCTGTTTTAACTTAACCTTAACCTGCAATGGAACCTCTTTTACTAGGAAT 180305
Qy 1029 gccacagcattgtctactactgattctctcagacttcatgcaaatgttttctcgtcag 1088
Db 180306 GCCACGATTTGCTACTACTGATTCGTTCTCAGACTTCTATGCAAAATGTTTCTCTGTCAG 180365
Qy 1089 atcaaacgatttaacttagaanaagctgtggagttacttagagactacttagttgat 1148
Db 180366 ATCAACAGTTTAAATCTAGAAAAGCTTGGAGTTTACTGTAGAGGACTACTTTAGTTGAT 180425
Qy 1149 gctgataaagtctactactgcaagaagctcgttttaataacagagagcgtctcacgta 1208
Db 180426 CTTGATAAATGGTCTACTTACTGCAAGAGCTGTTTAAATTAACGAGAGAGCTGCTCAGGTA 180485
Qy 1209 tctggtcagttcagattc 1226
Db 180486 TCTGTGTCAGTTTCAGATTC 180503
RESULT 7

CHTMOMP 1170 bp DNA BCT 26-APR-1993
LOCUS Chlamydia pneumoniae major outer membrane protein (MOMP) gene,
DEFINITION complete cds.
ACCESSION M64064 M34942 M64063
VERSION M64064.1 GI:144534
KEYWORDS major outer membrane protein.
SOURCE Chlamydia pneumoniae (strain IOL-207) DNA.
ORGANISM Chlamydia pneumoniae
REFERENCE 1 (bases 1 to 1170)
AUTHORS Carter, M.W., Al-Mahdawi, S.A.H., Giles, I.G., Trehan, J.D.,
Ward, M.E., and Clarke, I.N.
TITLE Nucleotide sequence and taxonomic value of the major outer membrane
protein gene of Chlamydia pneumoniae IOL-207.
JOURNAL J. Gen. Microbiol. 137, 465-475 (1991)
MEDLINE 91237311
FEATURES
Location/Qualifiers
1..1170
/organism="Chlamydia pneumoniae"
/strain="IOL-207"
/db_xref="taxon:83558"
1..69
/gene="MOMP"
1..1170
/gene="MOMP"
/codon_start=1
/transl_table=11
/product="major outer membrane protein"
/protein_id="AA23143.1"
/db_xref="GI:144535"
/translation="MKLLKSLALLSAFAAGSVGLQALPVGNPSDPSLLIDGTIWEGA
AGPCDPCATWCAISLRAGFYDYFDRLKVDAPKTFSGAKPTGSAANATTTAVD
RNPAYNKLHLDAEWFTRAGFIALNIWDREVFCTLGASNGYIRGNSTAFNLVGLGV
KGTIVANETPNVSLNSGVVLYDTSTFSVSGARGALWECGATLCAAEQYQASLPK
VEELNVCNVSQPSYKPKYKGVAPLPTDAGVATGTSATNVEHQVQASLSY
RLNSLPYIGVOWSRATFDADNIRIAOPKLPITAVLNLTANNPSLIGNATLSTDSFS
DFMIVSCQINKFKSRACGVTVGATLVDADKWSLTAEARLINEAAHVSGQFRF"
1..1170
/gene="MOMP"
mat_peptide 70..1167
/gene="MOMP"
/product="major outer membrane protein"
BASE COUNT 319 a 255 c 247 g 349 t
ORIGIN
Query Match 76.9%; Score 1096.4; DB 1; Length 1170;
Best Local Similarity 99.9%; Pred. No. 1.e-293;
Matches 1097; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 129 ttgccttagggaaacctctctgacagcttatttaattgattggtacaaatgggaaggt 188
Db 70 TTGCCTGTAGGGAACCTTCTGATCCAGCTTATTAAATGATGGTACAAATATGGAGGCT 129
Qy 189 gctgagagatccttgcagctcttgcctacttgcctgctgctgagcagctattgcttgcgt 248
Db 130 GCTGCAGGAGATCCTTGCCTACTTGCCTACTTGGTGCAGCGCTATTAGCTTACGTGCT 189
Qy 249 ggattttacggagactatgttttcgacctatcttaaaagtagatgcacctaacaattt 308
Db 190 GGATTTACGGAGACTATGTTTTCGACCGCTATCTTAAAAAGTAGATGCACCTAAACATTT 249
Qy 309 tctatggggaagcctactgagctgctgctgctgctgcaactactactcctccttagataga 368
Db 250 TCTATGGGAGCCAAAGCTACTGGATCCGCTGCTGCAAACTACTACTCTCCGTAGATAGA 309
Qy 369 cctaaccggtctacaataaagcatttacacgaagcagagtggttctactaatgcaggttc 428
Db 310 CCTAACCCGGCTTACAAATAGCATTTACAGGATGCAGAGTGGTGTCTACTTAAGCGCTTC 369
Qy 429 attgcttaaacatttggagcgttctgattttctgactttaggagcgttctaaggt 488

Db 370 ATTGCTTAACATTTGGGATCGCTTTGATGTTTTCTGTACTTTAGGAGCTTCTAATGCT 429

Qy 489 tacattaggaactctacagcttcaactcctgctgtttatttatttgcgaggttaaggtact 548

Db 430 TACATTAGAGAACTCTACAGCTTCAATCTCGTGGTGTATTCGGAGTAAAGGTACT 489

Qy 549 actgtaaatgcaaatgcaactacacacgctttctttaaagtaacgaggttctgaaacttac 608

Db 490 ACTGTAATGCAATGAACACTACCAACAGTTTCTTTAAGTAACGGAGTCTTGAACCTTAC 549

Qy 609 acagacacacctcttctcttctgagcgttagcgctctgagccttattggaatgcggttgc 668

Db 550 ACAGACACCTCTTCTTCTTGGAGCTGAGCGCTCGTGGAGCCTTATGGAATCGGTTCT 609

Qy 669 gcaactttggagctgaattcaaatatgcacagctcaaacctcaaaccttgaagtaaacactaat 728

Db 610 GCAACTTTGGAGCTGAATTCANATATGCACAGTCCAAACCTTAAGTTGAAAGAACTTAAT 569

Qy 729 gtgatctgaactgacgaattctctctgtataacacacacacacacacacacacacacacac 788

Db 670 GTGATCTGTAACTATCGCAATCTCTGTAAACAAACCAACCGCTATAAAGCGGTGCT 729

Qy 789 ttcccttgcacacagactgctgtagcaacagctacttggaacacacacacacacacacacac 848

Db 730 TTCCCTTCCCAACAGCCTGCGGTAGCAACAGCTTACTGGAAACAAAGTCTGCGACCATC 789

Qy 849 aattatcatgaatggcaagtagagcctctctctctctctctctctctctctctctctctctct 908

Db 790 AATTATCATGAATGGCAAGTAGGAGCCTCTCTATCTTACAGACTAAACTCTTTAGTGCCA 849

Qy 909 tacattggagtaacattgctgcgagcaacttttgcgtgataacacacacacacacacacacac 968

Db 850 TACATTGGAGTACAATGCTCGAGCACTTTTGTATGCTGATGAACACACGCGATGCTCAG 909

Qy 969 ccaaaactcctacagctgttttaacttaactgaactgaacacacacacacacacacacacacac 1028

Db 910 CCAAAACTACCTACAGCTGTTTAACTTAACCTGATGCAAGACCTTCTTTTACTAGGAAT 969

Qy 1029 gccacagcattgtctactactgattgcttctcagacacttcaatgcaaaattgttctctgcag 1088

Db 970 GCCACAGCATGCTCTACTACTGATGCTGCTCAGACTTTCATGCAAACTGTTCTCTGCAG 1029

Qy 1089 atcaacaagttaaatctagaaaagccttggaggttactgttaggagcactctttagttgat 1148

Db 1030 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 1089

Qy 1149 gctgataaatggtcacttactgcagaagctgtttaaattaaacgagagagcgtcactacgta 1208

Db 1090 GCTGATAATGTCTACTTACTGCAGAAAGCTGCTTTAATTAACGAGAGAGCTGCTCACGTA 1149

Qy 1209 tctggtcagttcagattc 1226

Db 1150 TCTGTCAGTTCAGATTC 1167

RESULT 8

CPMOMPX 1602 bp DNA BCT 21-MAY-1999

LOCUS Chlamydomonas reinhardtii gene for major outer membrane protein.

DEFINITION x72023

ACCESSION X72023.1 GI:313844

VERSION major outer membrane protein.

KEYWORDS Chlamydomonas reinhardtii.

SOURCE Chlamydomonas reinhardtii.

ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomonas.

REFERENCE 1 (bases 1 to 1602)

AUTHORS Girjes A.A., Carrick, F.N. and Lavin, M.F.

TITLE Remarkable sequence relatedness in the DNA encoding the major outer membrane protein of Chlamydia psittaci (koala type I) and Chlamydia pneumoniae

JOURNAL Gene 138 (1-2), 139-142 (1994)

MEDLINE 94171025

REFERENCE 2 (bases 1 to 1602)

Lavin, M.
Direct Submission
Submitted (22-APR-1993) M. Lavin, Professor of Molecular Oncology,
Queensland Institute of Medical Res, The Bancroft Centre, 300
Herston Road, Brisbane QLD 4029, AUSTRALIA
3 (bases 1 to 1602)
Lavin, M.
Direct Submission
Submitted (12-JUN-1993) M. Lavin, Professor of Molecular Oncology,
Queensland Institute of Medical Res, The Bancroft Centre, 300
Herston Road, Brisbane QLD 4029, AUSTRALIA
updates [2]
Location/Qualifiers
1. 1602
/organism="Chlamydomonas pneumoniae"
/db_xref="taxon:83558"
/note="biovar: koala; submitted as Chlamydia psittaci"
116^117
/citation=[2]
/replace="t"
239. 243
250. 1419
/codon_start=1
/transl_table=11
/product="major outer membrane protein"
/protein_id="CAA50906.1"
/db_xref="GI:313845"
/db_xref="SPTREMBL:O08085"
/translation="MKLLKSLLSAFAAGSVGSQALPVGNPSPDPSLLIDTIVEGA
AGDPCDCAWCDALSLRAGFYDVFDRILKVDAPKTFSMGAKPTSGATANYTTAVD
RNPAYNKLHDAEFTNAGFTALNIDWDFVCTLGASNGYIKGNSTAFNLVGLFGV
KGTSAANLNPVLSNGVELYTDTSFWSVGARGALWECGATLGAFOAQSKPK
VEELNVICNVAFQSVNPKGYKGVAPLPTDAGVATATGTSKATLNHFWQVGLSY
RLNSLPVYGVQWSRATEDADNIRIAOKPLPAVLNLTAWNPSSLGNTTTLATSDSFS
DFMQIVSCQINKFKSRKAGCTVGTATLVADADKWSLTAEARLLNERAAHVSGGFRP"
1450..1476
terminator 445 a 334 c 316 g 507 t
BASE COUNT
ORIGIN
Query Match 74.8%; Score 1066; DB 1; Length 1602;
Best Local Similarity 98.2%; Pred. No. 3.3e-285;
Matches 1078; Conservative 0; Mismatches 20; Indels 0; Gaps 0;
Qy 129 ttgcctgtgaggaacaccttctgacacgcttatttattgattgattgacacacacacacacacacac 188
Db 319 TTGCTGTAGGGAACCTTCTGATCCAGCTTATTATTAATGATGATGATGATGATGATGATGATGAT 378
Qy 189 gctgagggagacaccttgcgacaccttgcgacaccttgcgacaccttgcgacaccttgcgacacct 248
Db 379 GCTGCAGAGATCCTTTCGATCCTTTCGATCCTTTCGATCCTTTCGATCCTTTCGATCCTTTCGAT 438
Qy 249 ggaattttacgagac 308
Db 439 GGAATTTACGAGACTATGTTTTCGACCGCTATCTTTAAAGTAGATGATGATGATGATGATGATGAT 498
Qy 309 tctatgggagcaacaccttgcgacaccttgcgacaccttgcgacaccttgcgacaccttgcgacacct 368
Db 499 TCTATGGAGCAACACCTTCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGATCCTGAT 558
Qy 369 cctaac 428
Db 559 CCTAACCTCGCTACATAAGCATTTACAGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
Qy 429 attgccttaaac 488
Db 619 ATTGCCTTAACATTTGGAGATCCTTTGATGTTTCTGTACCTTAGAGCTTCTTAATGGT 678
Qy 489 tacattagaggaac 548
Db 679 TACATTAAAGGAACCTCTACAGCGTTCACTCGTGGTTTATTCGAGGTAAAGGTACT 738
Qy 549 actgtaaatgcaaatgcaactac 608

old_sequence
RBS
CDS
source

```

|||||
Db 739 TCTGTAGCTGCAATGAACCTACCAACGTTCTTTAAGTAATGAGTGTGTGAACCTAC 798
Qy 609 acagacacacctttctctcttgagagctgagcgcctctgagaccttatggagtcggtgt 668
Db 799 ACAGACACCTCTTCTCTTGAGGCTAGGCGCTCGTGAGCTTTATGGGAATGCGGTGT 858
Qy 669 gcaacttgggagctgaaattccaatatgcacagctccaaacccaaagttgaaagaaactta 728
Db 859 GCAACTTTGGGAGCTGAATTCATATGCACAGTCCAAACCTAAAGTTGAAGAATTAAT 918
Qy 729 gtgactgtaacgtatcgcaattctctgtataaacaacccaaagggctataaagggcttgc 788
Db 919 GTGATCTGTAACTAGCTCAATCTCTGTAAACAACCCAAAGGCTATAAGGCGGTGCT 978
Qy 789 ttcccttgcacaacagcgtgggttagcacagctactggaacaaagctctgcacaccatc 848
Db 979 TTCCCTCTTGCCAACAGCGCTGGCGTAGCACAGCTACTGGAACAAAGCTTGGCACCATC 1038
Qy 849 aattatcatgaatggcaagtaggagcctctctcttatctacagactaaaactctttagtcca 908
Db 1039 AATTATCATGAATGCAAGTAGGAGCTCTCTATCTTATAGACTAAACTCTTTAGTGCCA 1098
Qy 909 tacattggagtcacatgtctcgagcaacttttgcgtgataacatccgcatctgcag 968
Db 1099 TACATTGGAGTAAATGGTCTCGAGCAACTTTGTAGCTGTATGATACATCGCATGCTCAG 1158
Qy 969 ccaaaactcctacagctgttttaaaacttaactgaatgcagaaacccctctttactagaaat 1028
Db 1159 CAAAACACTACCTACAGCTGTTTAAACTTAACTGCTGATGAACCCCTCTTTACTAGGAAT 1218
Qy 1029 gccacagcattgtctactactgattgttcagaaagctgtgaggtactgtgagcactttagt 1088
Db 1219 ACGACAACATGGCTACTAGTATGCTGCTCAGACTTCATGCAAAATGTTTTCCTGTCAG 1278
Qy 1089 atcaacaagttaaaactagaaaagctgtgaggtactgtgaggtactgtttagttgat 1148
Db 1279 ATCAACAAGTTTAAATCTAGAAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGAT 1338
Qy 1149 gctgataaaatggtcactactgcagaagctcggttttaataacagagagagctgcacgta 1208
Db 1339 GCTGATAAATGGTCACTTACTGCGAAGCTCGTTTAAATTAACGAGAGAGCTGCTACGTA 1398
Qy 1209 tctggtcagttcagattc 1226
Db 1399 TCTGTCAGTTCAGATTC 1416

RESULT 9
CHTMOPEQ CHTMOPEQ 1170 bp DNA BCT 04-APR-1994
LOCUS Chlamydia pneumoniae major outer membrane protein gene, complete
DEFINITION cds.
ACCESSION L04982
VERSION L04982.1 GI:289840
KEYWORDS MOMP gene; major outer membrane protein.
SOURCE Chlamydia pneumoniae DNA.
ORGANISM Chlamydia pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
REFERENCE 1 (bases 1 to 1170)
AUTHORS Storey, C., Lusher, M., Yates, P. and Richmond, S.
TITLE Evidence for Chlamydia pneumoniae of non-human origin
JOURNAL J. Gen. Microbiol. 139, 2621-2626 (1993)
MEDLINE 94103736
FEATURES
    source
        1..1170
            /organism="Chlamydia pneumoniae"
            /specific_host="Equus caballus"
            /db_xref="taxon:83558"
    gene
        1..1170
            /gene="MOMP"
    sig_peptide
        1..69
            /gene="MOMP"

```

```

/note="putative"
1..1170
/gene="MOMP"
/note="precursor; putative"
/codon_start=1
/transl_table=11
/product="major outer membrane protein"
/protein_id="AA17397.1"
/db_xref="GI:289841"
/translation="MKKLLKSAALSAFAGSVGSLQALPVGNPSDPSLLIDTIIWEGA
AGDPDPCATCDIASLRAGFYGDIVFDRLIKIDAPKTFSMGAKPTGSATANYTTAVD
RNPAYNKHLYDAEWFTNAGFIALINWDRFDVFCILGASNGVYKNGSAFNLVGLFVG
KCTSVNANELPNVLSNGVIELYDTTFPWSYVARGALWECGATGFAQSKPK
VELNLVYIGVOMSRATFDADNIRIAQPKPLTALINLTAWNPSLLGSATAVSSDQFS
RLNSLPYIGVOMSRATFDADNIRIAQPKPLTALINLTAWNPSLLGSATAVSSDQFS
DPMQIVSCQINKEKSKKAGCTVGTATLVADKWSLTAEARLINERAHHISGGFR"
70..1167
/gene="MOMP"
/note="putative"
/product="major outer membrane protein"
312 a 246 c 251 g 361 t
BASE COUNT
ORIGIN
Query Match 69.9%; Score 997.2; DB 1; Length 1170;
Best Local Similarity 94.3%; Pred. No. 4.1e-266;
Matches 1035; Conservative 0; Mismatches 6; Gaps 0;
Qy 129 ttgctgtagggaaccccttctgacagcttataatgagtgatgataatggaaggt 188
Db 70 TTGCTGTAGGAACCCCTTCTGATCCAGCTTTATTAATGATGATGATGATGATGAT 129
Qy 189 gctcagagagatccttgcgactccttgcgactccttgcgactccttgcgactcct 248
Db 130 GCTCAGAGATCCTTGGATCCTTGGCTACTTGGTGGCAGCTTATAGCTTACGTCT 189
Qy 249 ggaatttcagagagactatgtttgcagcgtatctttaaagtagatgacccataacatt 308
Db 190 GGATTTTACGAGAGCTATGTTTTCGACCGTATCTTTTAAAGATAGATGCGCTTAAACATTT 249
Qy 309 tctatggagcgaacgctactggaatccgctgctgcaaaactactactgcgtagataga 368
Db 250 TCTATGGGAGCAACCTACAGGATCCGCTACTTGGAACTACTACTACCGCGCTAGATAGA 309
Qy 369 cctaaccggcgtacataaagaacattacacagatgcagagtggttcaataacgagcttc 428
Db 310 CCTAATCTCTTACATAAAGCATTTATACGATGAGATGCTTCTCCTTATTCAGGCTTC 369
Qy 429 attgcttaaacatttgggagcttgcgacttgcgacttgcgacttgcgacttgcgactt 488
Db 370 ATTGCTTTAAACATTTGGGATCGCTTGTATGTTTCTGTACCTGGAGCTTCTAATGCT 429
Qy 489 tacattagaggaaactctacagcttcaatctcgtttgtttatttcggaggttaaaagttact 548
Db 430 TACGTTAAAGGAAACTCTCGCGCTTCAATCTCGTTGTTTATTCGGAGTTTAAAGGCTACT 489
Qy 549 actgtaaatgcaaatgaaactacaaacgcttctttaaagtaacgaggtgttgaactttac 608
Db 490 TCTGTAAATGCAAAATGAACACTACAAACGTTCTTTTAAGTAATGGAGTTATCGAGCTCAT 549
Qy 609 acagacacacctcttctcttggagcgttagcgctcggtggagccttatgggaatcggttgt 568
Db 550 ACAGATACACTTTTCTGTTGGAGCGTAGGTGCTGCTGGAGCTTTATGGGAATTCGCGTGTCT 609
Qy 669 gcaacttgggagctgaaattccaatatgcacagctccaaacccaaagttgaaagaaactta 728
Db 610 GCCACTTTGGGAGCTGAATTTCCAAATATGCACAGCTCTAAACCTAAAGTTGAGAACCTAAT 669
Qy 729 gtgactgtaacgtatcgcaattctctgtataaacaacccaaagggctataaagggcttgc 788
Db 670 GTGATCTGTAACTAGCTCAATCTCTGTAAACAACCCAAAGGCTATAAGGCGGTGCT 729
Qy 789 ttcccttgcacaacagcgtgggttagcacagctactggaacaaagctctgcacaccatc 848

```

Db	730	TTCCCTTTGCCAACAGATGCTGGCGTAGTACACAGCTGCTGGAAACAAAGTCTGGACCATC	789
Qy	849	aattatcatgaatggcgaagtaggagcctctctatcttacagactaaacctctttagtgcca	908
Db	790	AATTATCATGAATGGCAGGTAGGAGCCCTCTATCTTATAGACTCAACTCTTTAGTGCCA	849
Qy	909	tcaattggagtacaatggctctcgagcaacttttgatgctgataacatcccgattgctcag	968
Db	850	TACATTGGAGTCCAAATGGTCTCGAGCAACTTTTGTATGCTGATAACATCCGATTGCTCAG	909
Qy	969	ccaaactactcacagctgttttaaacttaactcatcggaacccctcttttactaggaaat	1028
Db	910	CCAAAGTACTCAGCTATTTTAACTTAACTGCATGGAACCCCTCTTTACTAGGAGT	969
Qy	1029	gccacagcattgtctactactgattcgtcttcagacttcattgcgaatgttttcctgtcag	1088
Db	970	GCCACAGCTGTTCTTTCATCTGATCAATTCCTCAGATTCATGCAGATTGTGTCCTGTCAA	1029
Qy	1089	atcaacgaagttaaactctagaaaagcgtgtggagctactgttagagctacttttagtgat	1148
Db	1030	ATCAACAAGTGTAAATCTAGAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTTASTTGAT	1089
Qy	1149	gctgataaaggctcacttactcagaaagcgtcttgaattaaacagagagctgtcacgta	1208
Db	1090	GCTGATAAAGTGCATCTTACTGCGAAGAGCTCGTTTAATTACGAGAGAGCTGCTCACATA	1149
Qy	1209	tctggctcagttcagattc 1226	
Db	1150	TCTGTCAGTTCAGATTC 1167	
RESULT	10		
CHTOMPAAI			
LOCUS		999 bp DNA BCT 10-JUN-1999	
DEFINITION		Chlamydomophila pneumoniae outer membrane protein gene, partial cds.	
ACCESSION		M73038	
VERSION		M73038.1 GI:144566	
KEYWORDS			
SOURCE		Chlamydomophila pneumoniae.	
ORGANISM		Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.	
REFERENCE		1 (bases 1 to 999)	
AUTHORS		Kaltenboeck,B., Kousoulas,K.G. and Storz,J.	
TITLE		Structures of and allelic diversity and relationships among the major outer membrane protein (ompA) genes of the four chlamydial species	
JOURNAL		J. Bacteriol. 175, 487-502 (1993)	
MEDLINE		93123168	
FEATURES			
source		Location/Qualifiers	
		1..999	
		/organism="Chlamydomophila pneumoniae"	
		/strain="koala type I"	
		/isolate="koala"	
		/db_xref="taxon:83558"	
		/note="submitted as Chlamydia psittaci biovar: koala"	
		<1..>999	
		/note="ompA"	
		/codon_start=1	
		/transl_table=11	
		/product="outer membrane protein"	
		/protein_id="AA038210.1"	
		/db_xref="GI:5032262"	
		/translation="AGDCPCDPCATWCDAISLRAGFYGDYVDFILKVDAPKPTESMGAK	
		PGTSATNTVLTVDRENPAYNKHLDAENFTNAGVLSNGVVELYTDTSFMSVGARGALWECCGA	
		GNSTANVGLFVCGTSSVAANELPNVLSLNGVVELYTDTSFMSVGARGALWECCGA	
		TLGAEFYQAQSKPKVELINCVAFSNKPKGYGVAFPLEPTDAGVATATGKTAT	
		INTHWQVQASLSRYLNSLPVYIGVQWSRATFDADNRIAPQKPLTAVLMLTAEARLNPSL	
		LGNTHLPTLSDSFSDFMQIVSQCIQNFKSKKAGCTVGTATLVLDADKSLTAEARLIN"	
CDS			
		280 a 216 c 204 g 299 t	
BASE COUNT			
ORIGIN			

Query Match	67.8%	Score 967	DB 1	Length 999	
Best Local Similarity	98.0%	Pred. No. 1e-257			
Matches 979	Conservative 0	Mismatches 20	Indels 0	Gaps 0	
QY	192	gcaggagatcctgcgac	ccttcgcgtacttcggtgcacgc	tattagctacgtcgtcaga	251
DB	1	GCAGGAGATCCTTCGGATCCTTGGCTACTTGGTGGCGACGCTATTAGCTTTACGTGCTGGG	60		
QY	252	ttttcacggagactatg	tttttcgaccgctattcttaaaagt	tagatgcacctaaacattttct	311
DB	61	TTTTACGGAGACTATGTTTCGACCGTATCTTAAAGTAGATGCACCTTAAACATTTTCT	120		
QY	312	atgggagccaagccta	cgctgcgtgcgtgcacacta	ctactactgcgtagatagacct	371
DB	121	ATGGAGCCAAACCTACTGGATCGCTACTGCAACATATACTACTGCGGTAGATAGAGCT	180		
QY	372	aaccggcctacaataa	gaagcatttacacgatacgagtggtt	cactaaatgcaggcttcatt	431
DB	181	AACCCCTGCCATACATAAGCAITTTACAGATGCAGATGGTTTCACTTAATGCAGGCTTCATT	240		
QY	432	gccttaaacatttggg	atcgcgttcgtatgtttcttgc	tacttttagagctctctaaatggttac	491
DB	241	GCCTTAAACATTTGGGATCGCTTGTGATGTTTCTGTACCTTAGGAGCTTCTTAATGGTTAC	300		
QY	492	attagagaaactctac	gcgcttcaatctcgttcggttattc	gcgagtttaaaggctactact	551
DB	301	ATTAAGGAAACTCTACAGCGTTCAATCTCGTTGGTTTATTCGGAGTTAAAGGTACTTCT	360		
QY	552	gtaaatgcaaatgaac	taccacaaacggtttctttaa	gtaacgcgagttgttgaaactttacaca	611
DB	361	GTAGCTGCAAAATGAACATCCAAAGCGTTTCTTTAAGTAAATGGAGTTGTGAAACTCTACACA	420		
QY	612	gcaccccttctctga	gcgttaggcctcgtgcagccttat	gggaatgcggtgtgca	671
DB	421	GACACCTTCTTCTTGAGCGTAGGCGCTCGTGAGCTTTATGGGAATGCGGTGTGCA	480		
QY	672	acttgggagctgaatt	cctcaatgacagctccaaac	ctaaagttggaagaacttaatgtg	731
DB	481	ACTTTGGGAGTGAAATCCAAATATGCAGTGCCAAACCTAAAGTTGAAGAAGCTTAATGTG	540		
QY	732	atctgtaacgtatgc	gaattctctgtaaaaacccca	aggcgctatgaaggcggtcttc	791
DB	541	ATCTGTAACTGAGCTCAATCTCTGTAAACAAACCCAAAGGCGTATAAAGGCGTTGCTTTC	600		
QY	792	cccttgcacaacagc	gctggctagcaacagctact	ggaacaaagctgcgcacataaat	851
DB	601	CCCTTGCCACAGACGCTGGCGTAGCAACAGCTACTGGAAACAAAGCTGCGACCATCAAT	660		
QY	852	tatcatgaaatggca	agtaggcctctctatctt	cacagactaaaccttttagtgccatac	911
DB	661	TATCATGAATGGCAAGTAGGAGCGCTCTATCTTATAGACTTAAACTCTTTTAGTGCCATAC	720		
QY	912	attgagatcaaatg	ctcgaacacitttgatctga	ataacatccgattgctcagcca	971
DB	721	ATTGGAGTACAAATGGTCTCGAGCAACTTTTGATGCTGATAACATCCGCGATTGCTCAGCCA	780		
QY	972	aaactacctacagc	gttttaaacttaactgc	atggaacccctctcttactaggagaaatgcc	1031
DB	781	AAACTACCTACAGCTGTTTTAACTTAACATGCATGGAACCCCTCTCTTTACTAGGAAATACG	840		
QY	1032	acagcattgtctac	tactgattcgttctcaga	acttcattgcaaatgtttctcgtcagatc	1091
DB	841	ACAACATTTGCCATCTAGTAGTTCGTCTCAGACTTCATGCAAAATGGTTTCTGTGCAGATC	900		
QY	1092	aacaagtttaaatc	tagaagaagcttggaggtt	actctgttaggagctacttttagttagtgc	1151
DB	901	AACAAGTTTAAATCTAGAAAGCTTGTGGAGTTACTGTAGGAGCTACTTTAGTTGATGCT	960		
QY	1152	gataaatggtcact	taactgcagaagcgtcgtt	taattaac	1190
DB	961	GATAAATGGTCACTTACTGCAAGAGCTCGTTTAATTAAC	999		

QY	600	gaactttacacagacacacctcttcttgagcgtagcgctgtagagcctttaggaa	659
DB	241	GAACCTTTACACAGACACCTCTTCTTCTTGAGCGTAGGCGCTCTGAGCCCTTATGGGAA	300
QY	660	tgcggtgtgcaactttggagctgaattccaatatgcacagtcacacccaaagttaa	719
DB	301	TGCGGTGTGCAACTTTGGAGCTGAATTCMAATATGCACAGTCCAAAGCTAAAGTTGAA	360
QY	720	gaacttaattgctgtaacgtatcgcaattctctgtaaacacacccaaaggctataa	779
DB	361	GAACCTTAATGTGATCTGTAACGTATCGCAATCTCTGTAACAAACCCCAAGGCTATATA	420
QY	780	ggcgtgtcttcccttgcacacagcgtggcgtagcaacagctactggaacaaagtct	839
DB	421	GGCGTTGCTTCCCTTGCCACAGAGCTGGCGTAGCAACAGCTACTGGAACAAAGTCT	480
QY	840	gcgaccatcaattatcatgaatggcaagtaggagcctctctattctacagactaaact	899
DB	481	CGCACCATCAATTATCATGAATGCAAGTAGGAGCCTCTCTATCTTACAGACTAACTCT	540
QY	900	ttagtgcacacattgagtagacaatggtctcgagcaacttttgatgctgataacatccg	959
DB	541	TTAGTGCATACATTGGAGTACATGCTCTCGAGCAACTTTTGATGCTGATAACATCCG	600
QY	960	attgctcagcaaaactacacagcgtgttttaaaacttaactgacagaccccttttta	1019
DB	601	ATTGCTCAGCCAAACTACCTACAGCTGTTTTTAACTTAACATGATGCAACCCCTTTCTTA	660
QY	1020	ctaggaataatg 1029	
DB	661	CTAGGAATAATG 670	
RESULT	12		
AF131229			
LOCUS	670 bp	DNA	BCT 01-JUN-1999
DEFINITION	Chlamydomophila pneumoniae mutant major outer membrane protein (MOMP)		
ACCESSION	AF131229		
VERSION	AF131229		
KEYWORDS	AF131229.1	GI:4928267	
SOURCE	Chlamydomophila pneumoniae.		
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.		
REFERENCE	1 (bases 1 to 670)		
AUTHORS	Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.		
TITLE	Presence of viable Chlamydia pneumoniae in fetal calf serum and in epithelial-derived cell lines		
JOURNAL	Unpublished		
REFERENCE	2 (bases 1 to 670)		
AUTHORS	Tharp,A.C., Mitchell,W.M., Stratton,C.W. and Ding,L.-M.		
TITLE	Direct Submission		
JOURNAL	Submitted (25-FEB-1999) Pathology, Vanderbilt University, C-3321 Medical Center North, Nashville, TN 37232, USA.		
FEATURES	Location/Qualifiers		
source	1..670		
	/organism="Chlamydomophila pneumoniae"		
	/db_xref="taxon:83558"		
	/note="synonym: Chlamydia pneumoniae"		
	isolated from commercial fetal calf serum"		
gene	<1..>670		
	/gene="MOMP"		
CDS	<1..>670		
	/gene="MOMP"		
	/note="similar to MOMP gene sequences of Chlamydia pneumoniae presented in GenBank Accession Numbers M64064, M34942, and M64063"		
	/codon_start=1		
	/product="mutant major outer membrane protein"		
	/protein_id="AAD33511.1"		
	/db_xref="GI:4928270"		
	/translation="VDRENPAYNKHLHDAEWFNAGFTALNIWDFVDFCTLGASNGY IRNSTAEFLVGLFGVGTNNANLPSVLSNGVLYDTISFWSVGAGALWECC CATLGAETQYAGSKPELVANVCVQSVNPKYGVAPLPTDAGVATATGTS ATINHEWQVGSASLYRLNSLVPYIGVQWSRATEDADNRIRIAQPKLPTAVLNLTAWN		
variation	SLIGN"		
	63		
	/gene="MOMP"		
	/note="difference from wild type, silent mutation"		
	/replaces="a"		
BASE COUNT	191 a	147 c	138 g
ORIGIN	194 t		
Query Match	46.9%; Score 668.4; DB 1; Length 670;		
Best Local Similarity	99.9%; Pred. No. 9e-175;		
Matches	669; Conservative	0; Mismatches	1; Indels
	0; Gaps		
0;			
QY	360	gtacatagacctaacccgctcaataagcattacacagtcagagtggttcaactaat	419
DB	1	GTAGATAGACTAACCCGGCTACATACATATTACAGATGCAGAGTGGTTCACTAAT	60
QY	420	gcaggtcttcattgcttaaacattggagtcgcttgatgtttctgtacttttagagct	479
DB	61	GCCGCTTCATTGCCTTAACATTTGGATCGCTTGTGATGTTTCTGTACTTTAGGAGCT	120
QY	480	tctaatggttacatagagaaactctacagcgttcaatctctgttttattcgaggt	539
DB	121	TCTAATGGTTACATTAGAGAACTCTACAGCGTTCATCTCGTGTGTTTATTCCGGAGTT	180
QY	540	aaaggctactctgtaaatgcaatgaactaccacacgctttctttaacgacgagttggt	599
DB	181	AAAGGTACTACTGTAAATGCAATGAATGAATCAACCAACCTTCTTTAAGTAACGGAGTTGTT	240

IRGNSTANLVGLFVKGTTVNANELPNVLSNGVVELYDTDFSWSGARGALWECG
CATLGAQFQAQSPKVEELNVICNVQFVSNKPKYKGVAFPLPTDAGVATATGTS
VTINHEWQVGCASLYSLNVLPIGVQWSRATEDADNIRIAQPKLPVAVLNLTWNE
SLLEN"

variation

/gene="MOMP"

/note="difference from wild type; in wild type encodes

pro"

/replace="c"

29

/gene="MOMP"

/note="difference from wild type; in wild type encodes

lys"

/replace="a"

482

/gene="MOMP"

/note="difference from wild type; in wild type encodes

ala"

/replace="c"

191 a 144 c 139 g 196 t

BASE COUNT
ORIGIN

Query Match 46.6%; Score 665.2; DB 1; Length 670;

Best Local Similarity 99.6%; Pred. NO. 7e-174;

Matches 667; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 360 gtatagacctaaccgcccacataaagcattacacgacgagaggttcacaaat 419

DB 1 GTAGATAGACCTAACCGCCCTACAAATATGCAATTTACAGATGCGAGGTTCACATAAT 60

QY 420 cgaggttcattgccttaaacatttgggacgtcttgatgttttctgtacttttagagct 479

DB 61 CGAGGCTTCATTGCGCTTAACATTTGGGATCGCTTTGATGTTTCTGTACTTTTAGAGCT 120

QY 480 tctaattggttacattagaggaacacctacagcgttcaatctcgttgggtttatcgagtt 539

DB 121 TCTAATGTTTACATTAGAGGAACCTCTACAGCGTTCAATCTCGTTGGTTTATTCGGAGTT 180

QY 540 aaaggtactactgttaaatgcaaatgaactacacacgcttcttcttaagtaacgaggtgtt 599

DB 181 AAAGGTACTACTGTAAATGCAANTGAATCACTACCAACGTTCTTTAAGTAACGAGTTGTT 240

QY 600 gaacttacacagacacacctcttctcttcttcttcttcttcttcttcttcttcttcttctt 659

DB 241 GAACTTACACAGACACCTCTTCTCTTGGAGCGTAGGCGCTCGTGGAGCCTTATGGGAA 300

QY 660 tgcggttgcaactttggagctgaattccaatgacacgctccttcttcttcttcttcttcttct 719

DB 301 TCGGTTGTGCAACTTTGGAGCTGAATTCCAATATGCACAGTCCCAACCTTAAAGTTGAA 360

QY 720 gaacttaattgtatgtatacgtatcgcaattctcttcttcttcttcttcttcttcttcttctt 779

DB 361 GAACTTAATGTATGTAACTGTAACGTATCGCAATTCCTGCTAAACAAACCCCAAGGCTATAA 420

QY 780 ggcgtgttctcccttcccaacagacgctggcgtagaacaagctactggaacaagctct 839

DB 421 GCGGTTGTTTCCCTTCCCAACAGACGCTGGCGTAGCAACAGCTACTGGAACAAAGTCT 480

QY 840 gcgaccatcaattatcatgaatggcaagttaggagcctctcttcttcttcttcttcttcttct 899

DB 481 GTGACCATCAATTATCATGAAATGGCAAGTAGGAGGCTCTCTATCTTACAGACTTAACTCT 540

QY 900 ttatgcccatacattggagtaacattggtctcgagcaactttttgatgctgataacatccgc 959

DB 541 TTAGTGCCATACATTGGAGTACAAATGTTCTCGAGCAACTTTTGTGCTGATAAATCCGC 600

QY 960 attgctcagcaaaaactacctacagctgttttaacttaacttaacttaacttaacttaacttta 1019

DB 601 ATTGCTCAGCAAAAACCTACCTACAGCTGTTTAAACTTAACTGATGCAACCCCTCTTTA 660

QY 1020 ctaggaaatg 1029

DB 661 CTAGAAATG 670

RESULT 13

AF269259

LOCUS

DEFINITION

AF269259

ACCESSION

AF269259.1

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

1..1425

/organism="Chlamydomophila psittaci"

/strain="VS225"

/specific_host="orange-fronted parakeet; Aratinga

canicularis"

/db_xref="taxon:83554"

/country="USA; Laredo, Texas"

/note="isolated from swab material submitted to APHIS in

August 1991 from a bird purchased in Laredo, TX USA;

generated with the primers CTU (5' ATGAAAAAAGCTTTGAATCGG

3') and TGLY (5' GGCTACAGCTCTACCATGTA 3')."

<1..1167

/gene="ompA"

<1..57

/gene="ompA"

<1..1167

/gene="ompA"

/note="porin; MOMP"

/codon_start=1

/transl_table=11

/product="major outer membrane protein"

/protein_id="AAK00240.1"

/db_xref="GI:12642455"

/translation="LLKSAALLFAATGSALSQAALPVGNPAEPSSLIDGTMMWEGASGDP

CDPCATWCDAISIRAGYGDYVDFVRLKVDVNTKISGMAAPTSAAADYKTPTRPN

IAYCKHLQDAEMFTNAFLALNINWDRDIFCTLGASNGYFKASSAANLVGLVKGKT

SVAADLPNVGITQGIIVEFYDTTFSVSGARGALWEGCGCATLGEFOYQASNPALIEH

LNVLSSPQFVVVHRPRGKGTGSNFPPLPTAGTDGATDTKSLTLKHYHMQVGLALSFR

LNMLVPIYGVNWRATFDADSIIRAKPLAAVLNLTNTNPTLLGEATLDAENKFCR

FLQASTQINKKSKRACGAVGATLIDADKWSITGEARLINERAAHMAOFRF"

58..1164

/gene="ompA"

/product="major outer membrane protein"

1395..>1425

/product="tRNA-Gly"

complement(1406..1425)

BASE COUNT 394 a 313 c 294 g 424 t

ORIGIN

Query Match 42.5%; Score 606.6; DB 1; Length 1425;

Best Local Similarity 73.1%; Pred. NO. 1.5e-157;

Matches 809; Conservative 0; Mismatches 289; Indels 9; Gaps 2;

QY 129 ttgctgtagggaaccccttctgtatccaagcttataatgagtggtacataatgggaagct 188

|||||

Db	58	TTGCGCTAGGGAAACCCAGCTGAACCAAGTTTATTAAATGATGGCACTATGTGGGAAGGT	117
Qy	189	qctgcaggagatccttgcgatccttgcgctacttggtgcagcctattagcttactgct	248
Db	118	GCTTCAGAGATCCTTCGGATCCTTGGCTTACTTGGTGTGACGCCATTAGCATCCGGCA	177
Qy	249	ggattttacggagactatgttttgcgcgtatctttaaagttagatgcacctaataacatt-	307
Db	178	GGATACTACGGAGATTATGTTTTCGATCGTGTAATAAAGTTGATGTGAATAAACAATATC	237
Qy	308	--ttctatggagccaagcctactggatccgctgctgcaactatactactgcgcgtagat	365
Db	238	AGCGGTATFGGTGCGAGTCTTACAGGAAGGACGACGCCGATTACAAACACTCTACAGAT	297
Qy	366	agacctaaccggcctacaataaagcatttacacgatcgagatggttccataatgcaggc	425
Db	298	AGACCCACATCGCTTATGGCAACACTTGCAGACGCTGAGTGGTTTTCAGGAATGCACT	357
Qy	426	ttcattgcttaaacattgggactgcttgatgttttctgtaactttagagcttctaatt	485
Db	358	TTCTCTCGATTAAATATCTGGGATCGCTTTCGATATTTCTGCACATTAGGTGCTTCCAAAT	417
Qy	486	ggttacattagagaaacctctacagcgttcaactcctggtgtttattcggaggttaaaaggt	545
Db	418	GGGTACTTCAAGACTAGTCTCTGCTGCAATCAACCTCGTTGCTTTCATTGGTGTATAAGGA	477
Qy	546	actactgtaaatgcaaatgaaactccaaacgctttctttaagtaacggagtttgttgaactt	605
Db	478	ACCTCGGTAGCAGCTGATCAACTCCAAACCTGAGGCATCACTCAAGTATTGTTGAGTTT	537
Qy	606	tacacagacacctcttctctggagctgagcgctgctggagccttatggaaatgcggt	665
Db	538	TACACAGATACAACTTCTCTGGAGCGTAGGTGCACGCTGGTCTTATGGAAATGTGCT	597
Qy	666	tgtgcaactttggagctgaaattccaaatatgcacagtcctcaaacctaaagttgaagaactt	725
Db	598	TGTGCAACTTTAGGAGCTGAATTCAGCATGTCTCAATCTAATCTCTTAAATTTGAAATGCTG	657
Qy	726	aatgtgatctgtaagctatcgcaattctctgtaaaacacccaaggcgctataa-----a	779
Db	658	AATGTAAATCTCCAGGCCAACAACTTTGTAGTTTCACAGCCTTAGAGGATACAGGGAACA	717
Qy	780	ggcgttgcttcccttgcacacagcgtgctgtagcaacagctactggaacaaagtct	839
Db	718	GGATCGAACTTTCCCTTTACCTCTAACAGCTGGTACAGATGGTGTACAGATACTAAATCT	777
Qy	840	gogaccaataattatcatgaatggaagtagagcctctctatctacagactaaactct	899
Db	778	GCAACACTCAAAATATCATGAATGGCAAGTTGGTTTAGCGCTCTCTTACAGATTGAACATG	837
Qy	900	ttagtgcatacattggagtcaaatggtctcgagcaacttttgatgctgatacaatccgc	959
Db	838	CTTGTGTCCTTACATTGGCGTAACTGGTCAAGAGCAACTTTTGTATGCTGACTCTATCCGC	897
Qy	960	attgctcagccaataactacagctgcttttaaaccttaactgcatggaaccccttcttta	1019
Db	898	ATCGCTCAACCTTAAATTAGCGCTGCTGTTTGAAGTTGACCACATGGAAACCAACTCTT	957
Qy	1020	ctaggaatgcacagcattgctactactgattgcttccagacttcatgcaaatgtt	1079
Db	958	TTAGGGGAAGCTACAGCTTTTATAGTGTAGCAACAAATCTCGGACCTTTTACAAATCGCT	1017
Qy	1080	tcctgtcagatcaacaagtttaacttagaaaaagcttggtgagttactgttagagctact	1139
Db	1018	TCGATTTCAGATCAACAAATGAATCTPAGAAAAAGCTTGTGGTGTAGCTGTTGTGGCAACG	1077
Qy	1140	ttaagttagctgataaattgctcattactgcagagctcggtttaaataacgagagagct	1199
Db	1078	TTAATGACCGCTGACAAATGGTCANTCACTGGTGAACGACGCTTAATCAATGAAGAGCT	1137
Qy	1200	gctcacgtatctgctcagttcagattc	1226
Db	1138	GCTTACATGAATGCTCAATTGAGTTTC	1164

RESULT	14
AF272945	1170 bp DNA BCT 25-JAN-2001
LOCUS	Chlamydomophila abortus major outer membrane protein precursor (Ompl)
DEFINITION	gene, complete cds.
ACCESSION	AF272945
VERSION	AF272945.1 GI:12483891
SOURCE	Chlamydomophila abortus
KEYWORDS	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
ORGANISM	1 (bases 1 to 1170)
REFERENCE	Vretou,E., Psarrou,E., Kaisar,M., Vlisidou,I., Saiti-Montesanto,V. and Longbottom,D.
AUTHORS	Identification of protective epitopes by sequencing of the major outer membrane protein gene of a variant strain of Chlamydia psittaci serotype 1 Infect. Immun. 69 (1), 607-612 (2001)
TITLE	20569239
JOURNAL	11119563
MEDLINE	2 (bases 1 to 1170)
PUBLISHED	Longbottom,D. and Vretou,E.
REFERENCE	Direct Submission
AUTHORS	Submitted (29-MAY-2000) Moredun Research Institute, Bush Loan,
TITLE	Penicuik EH26 0PZ, UK
JOURNAL	Location/Qualifiers
FEATURES	1..1170
source	/organism="Chlamydomophila abortus"
:	/isolate="LJG"
:	/specific_host="Capra aegagrus hircus"
:	/db_xref="taxon:83555"
:	/country="Greece"
sig_peptide	1..66
:	/gene="Ompl"
CDS	1..1170
:	/gene="Ompl"
:	/note="MOMP; porin"
:	/codon_start=1
:	/transl_table=11
:	/product="major outer membrane protein precursor"
:	/protein_id="AAG53881.1"
:	/db_xref="GI:12483892"
:	/translation="MKKLLKALLFRATGSALSLSQALPYCNPAEPSILLIDGTWEGAS RDCPCSTCDAISIRAGYGVDFVRLKVDNKTITGMGAVPPTGTAADYKTFTD GNPIAYGKHLDIAEWFETNAFALAINWDREIDFTLGASNGYFKASAAPNLVLGG KSSVAADQLPNVGITGCIVEFYDTTFWSVGARGALWECCGATLGAEPYQAQSPK TEMLNVSDQAPGVVHRPRGYKGTFPLTAGTDQATDKSATIKYHEWQGLAIS RLNMVPIYGVNNSRAFDDAIRIAPOKIAAAVLNTWNPTLLGEATTLDTNKKFA DLFIQIASIQINKMSRKACGVAVGATLIDAKSKSIUTGEARLINERAHHMAAQPF"
gene	1..1170
:	/gene="Ompl"
mat_peptide	67..1167
:	/gene="Ompl"
:	/product="major outer membrane protein"
BASE COUNT	322 a 263 c 255 g 330 t
ORIGIN	
:	Query Match 42.1%; Score 600.2; DB 1; Length 1170;
:	Best Local Similarity 72.2%; Pred. No. 8.6e-156;
:	Matches 795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;
QY	129 ttgctgttgagggaacctctatccaagccttattaatgatgggtacatatggaaagt 188
Db	67 TTGCCCTGTAGGGAACCACCTCAACCAAGCTTTTATTAACTCGATGGCACATGTGGGAAGT 126
QY	189 gctgcaggagatccttcgcactcttcgcctacttgtgcgcagcctattagcttagctgt 248
Db	127 GCTTCAGTGTAUCCCTTGCATCCTTGCTACTTGTGTGTGATCATCATGCATCCGGCA 186
QY	249 ggattttacggagaactatgttttccaccgcctatcttaaaagtgatgacgtcaactaaacattt 308

```

187 GGATACACGAGATATGTTTCGATCGTGTATTAAGAGTGTGATGAATAAACTATC 246
QY 309 tc---tatggagcaagcactgagtcggtgcaaacataactactactccgtagat 365
Db 247 ACCGGCATGGGTGCGATGCTTACAGGAACCGGAGCAGCTGTATACAAAACCTCCTACGGAT 306
QY 366 agacctaaacccgctacataaagcatattacacgagtgagagtggttcaactaatgcagc 425
Db 307 AGACCAACATCGCTTACGGCAACACATTTACAGAGCCGCGATGGTTACCAATCGAGCT 366
QY 426 ttcatggccttaaacatttggagtcgcttgatgttttctgactttaggagcttcta 485
Db 367 TTCTCGCATTAATATCTGGGATCGCTTGTATTTCTGCACATTTAGCGCTTCTAAT 426
QY 486 ggttcattagaggaactacacgcttcaactcgttggttttattcgaggttaaggt 545
Db 427 GGGTACTTCAAGCTAGTTCTCGGCATTTCAACCTCGTGGTGTGATGGTGTAAAGGA 486
QY 546 actactgtaaatgcaaatgaactacacaaagctttctttaagtaacgaggttggaaactt 605
Db 487 TCCTCGTAGCAGCTGATCAGCTTCCCAATGTAGGCATCACCAAGGAATCGTTGAATTT 546
QY 606 tacacagacacactcttcttcttggagcgtgagcgtctgagccttattgggaatgcggt 665
Db 547 TATACAGATACAACTATCTTCTGGAGTGTAGTGCACGCGAGCTTTATGGAGTGTGT 606
QY 666 tgtgcaacttggagctgaattcacaatatgcacagtcacaaacccctaaagtgaagaactt 725
Db 607 TGTGCACTTTAGGAGCAGAGTTCCAAATGAGCTGCTTATCTTAAATTTGAATGTTG 666
QY 726 aatgtgactgttaacgtatcgcaattctctgttaacaaaccccaagggctataaagcgctt 785
Db 667 AATGTAGTCTCAGGCCAGCAAAATTTGTGGTTCACAAAGCTTAGAGGATACAAAGGAACA 726
QY 786 gcttcccttgcacacagacgctgctgtagcaacagctactggaacaaagctcgcgacc 845
Db 727 GCAATTCCTTACCTACAGCTGGTACTGATGACGGCAACTGACACTAAGTCGGCTACA 786
QY 846 atcaattatcatgaatggcaagtaggagcctctctatcttcaagactaaactctttagtg 905
Db 787 ATTAATACACGAATGGCAAGTTGGTTTACGCTCTCTTATCGATGAACATGCTTGT 846
QY 906 caacattggagtagcaaatgctgcgagcaactttttagctgatacaactcgcgactgt 965
Db 847 CCTTACATTTGGGCTAAACTGGTACAGCAAACTTTTGTATGCTGAGCTATCGGCATCGCT 906
QY 966 cagcaaaactcctcacagctgttttaaaacttaactgcatgaaacccctcttctactaga 1025
Db 907 CAACCTAATTAGCTGCTGCTGTATTAACCTGACCAATGACCAATGACCAACCCCTTTAGGG 966
QY 1026 aatgcaacagcattgtctactactgattgcttctcagacttcatgcaaatgttctcgt 1085
Db 967 GAAGCTACAACTTTAGATTAACCAAAATTCGCTGACTTCTTGCAAATTTGCTCGATT 1026
QY 1086 cagatcaacaagtttaactttagaagaactgttgagttactgtaggagactttagtt 1145
Db 1027 CAGATCAACAAATGAAGTCTAGAAAAGCTTGTGGTGTAGCTGTGGTGCACACGTTAATC 1086
QY 1146 gatgctgataaagtgtcactactcagagagctcgttttaataacgagagagcgtcctac 1205
Db 1087 GACGCTGACAAATGGTCAATCACTGGTGAAGCAGCAGCTTAATCAATGAAGAGCCGCTCAC 1146
QY 1206 gtaactggtcagttcagattc 1226
Db 1147 ATGAATGCTCAATTCAGATTTC 1167

```

RESULT 15
 CHOMPAD 1261 bp DNA BCT 07-MAY-1999
 LOCUS Chlamydomophila abortus strain B577 major outer membrane protein
 DEFINITION precursor (ompA) gene, complete cds.

```

M73036
M73036.1 GI:144564
Chlamydomophila abortus.
Chlamydomophila abortus
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE
1 (bases 1 to 1261)
AUTHORS
Kaltenboeck, B., Kousoulas, K.G. and Storz, J.
TITLE
Structures of and allelic diversity and relationships among the
major outer membrane protein (ompA) genes of the four chlamydial
species
JOURNAL
J. Bacteriol. 175, 487-502 (1993)
MEDLINE
93123168
FEATURES
Location/Qualifiers
1..1261
/organism="Chlamydomophila abortus"
/strain="B577"
/specific_host="Ovis aries"
/db_xref="taxon:83555"
/note="submitted as Chlamydia psittaci"
80..1246
/gene="ompA"
80..1249
/codon_start=1
/transl_table=11
/product="major outer membrane protein precursor"
/protein_id="AAD29102.1"
/db_xref="GI:4757612"
/translation="MKKLSKALLFAATGSAALSQAALPVGNPAEPSSLIDGTWEGAS
GDPDPCSTWCDATSIKAGYDYDFRVLVKNTITGMGAVPTGTAANAKTPTD
RNTAYGKHLQDAEWFNTAAFLINIMDFDICTLGASNGYFKASSAFLNLVLIGV
KPSSTAADQLPNVIGTQIGVFYDTTFSWSVGARGALWECGATLGAEFQVQSNPK
IEMLNVSPPAQFVHKPRGYTAPLPLTAGTDOATDTKSATKYHEWQVGLALSY
RLMLVPYISVNSRATDADAIKTAOPKLAALNLNTWNPITLAGEATLDTSKFA
DFLOIASIQINKMKSRKACGVAVGATLIDADKWSITGEARLINERAHMNAQFRF"
BASE COUNT 359 a 275 c 277 g 350 t
ORIGIN

```

```

Query Match 42.1%; Score 600.2; DB 1; Length 1261;
Best Local Similarity 72.2%; Pred No. 8.6e-156;
Matches 795; Conservative 0; Mismatches 303; Indels 3; Gaps 1;
QY 129 ttgccttagggaaacccctctgacaaacttattatgagtgagtaacatgggaaggt 188
Db 146 TTGCCTGTAGGGAACCCAGCTGAACCAAGTTTATTATTCATCGATGGCACTATGTGGGAAGGT 205
QY 189 gctgcagagatccttgcgactcctgcgacttggcgagcgtattagcttaagctgt 248
Db 206 GCTTCAGGTGATCTTCGATCTTGTCTACTTGTGTGATGCTATCAGCATCCGCGCA 265
QY 249 ggattttacggagactatgttttcgacgctatcttataaagtagatcacctaaacattt 308
Db 266 GGATACACGAGATATGTTTCGATCTGTATTAAGTTCATGTAATAAACTATC 325
QY 309 tc---tatggagcaagcactgagtcgctgctgcaaacataactactactcgcgtagat 365
Db 326 ACCGGCATGGGTGCGATGCTTCTACAGGAACCGGAGCAGCTTAATACAAAACCTCCTACGGAT 385
QY 366 agacctaaacccgctacataaagcatattacacgagtgagagtggttcaactaatgcagc 425
Db 386 AGACCAACATCGCTTACGGCAACACATTTACAGAGCCGCGATGGTTACCAATCGAGCT 445
QY 426 ttcatggccttaaacatttggagtcgcttgatgttttctgactttaggagcttcta 485
Db 446 TTCTCGCATTAATATCTGGGATCGCTTGTATTTCTGCACATTTAGCGCTTCTAAT 505
QY 486 ggttcattagaggaactcctacacgcttcaactcgttggtttattcgaggttaaggt 545
Db 506 GGGTACTTCAAGCTAGTTCTCGGCATTTCAACCTCGTGGTGTGATGGTGTAAAGGA 565
QY 546 actactgtaaatgcaaatgaactacacaaagctttctttaagtaacgaggttggaaactt 605

```

Fri Feb 8 08:44:50 2002

Db 566 TCCCTCATAGCAGCTGATCAGCTTCCCAATGTAGGCATCACTCAAGGAATCGTTGAATTT 625
Qy 606 tacacagacaccttcttctcttgagcgtagcgctctgtagccttatgggaatgcggt 665
Db 626 TATACAGATACAACATCTCTTGGAGTGTAGGTGCACGGGAGCTTTATGGAGTGTGGT 685
Qy 666 tgtgcaactttgggagctgaattccaatatgcacagtcaccaaacctaaagtgaagaactt 725
Db 686 TGTGGACTTTAGGAGCAGAGTTCCAATACGCTCAGTCTAATCCTAAATTTGAATGTTG 745
Qy 726 aatgtgattctgtaacgtatcgcaattctctgtataaacaacccaaaggcctataaaggcgtt 785
Db 746 RATGTAGTCTCCAGCCAGCACAATTTGTGTTTACAGCCTAGAGGATACAGGGGAACA 805
Qy 786 gctttcccttgcgaacagcgctggcgtagcaacagctactggaacaaagtctgcgacc 845
Db 806 GCATTTCCCTTTACCTCTAACAGCTGGTACTGATCAGGCAACTGACACTAAGTCGGCTACA 865
Qy 846 atcaattatcatgaatggcaagtaggagcctctctctattacagactaaactctttagtg 905
Db 866 ATTAATACACGAATGGCAAGTTGGTTTAGCGCTCTCTTTATCGATTGAACATGCTTGT 925
Qy 906 ccatacattggagtagcaatggctctcgagcaactttttagctgataacatcgcgattgct 965
Db 926 CCTTACATTAGCGTAAACTGGTACAGCAGCACTTTTGATGCTGAGGCTATCGGCATCGCT 985
Qy 966 cagccaaaactacacagctgttttaaaacttaactgcatggaacccctcttactagga 1025
Db 986 CAACCTAAATTAGCTGCTGCTGTGTTAACTTGACACATGGAGCCCAACCCCTTTTAGGA 1045
Qy 1026 aatgcacagcattgtctactactgattgcttctcagacttcatgcaaatgtttccctgt 1085
Db 1046 GAAGCTACAGCTTTAGATACTAGCAACAATTCGCTGACTTCTTGCAAAATGCTTCGATT 1105
Qy 1086 cagatcaacaagtttaaatctagaaaagctgtggagttactgtgaggtactttagtt 1145
Db 1106 CAGATCAACAAAATGAAGTCTAGAAAAGCTTGTGGTGTAGTGTGGTGCACGTTAATC 1165
Qy 1146 gatgtgataaattgggtcacttactgcagagctcgttttaattacagagagagctgtcac 1205
Db 1166 GAGCGTGACAAAATGGTCAATCATCGTGAAGCAGCGCTTAATCAATGAAGAGCCGCTCAC 1225
Qy 1206 gtatctgggtcagttcagattc 1226
Db 1226 ATGAATGCTCAATTCAGATTC 1246

Search completed: February 7, 2002, 19:37:19
Job time: 13753 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 20:08:36 ; Search time 685.35 Seconds
(without alignments)
1783.827 Million cell updates/sec

Title: US-09-391-606-12

Perfect score: 1426

Sequence: 1 tgcagctactctgtctgc.....tcctgacctgaaggtgc 1426

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	600.2	42.1	1261	20 AAX25047	Chlamydia psittaci
2	526.4	36.9	1660	20 AAX25048	Chlamydia psittaci
3	523.2	36.7	1209	20 AAX25046	Chlamydia psittaci
4	514.6	36.1	1578	21 AAA64764	C. pneumoniae sero
5	514.6	36.1	1578	22 AAH56267	Chlamydia trachoma
6	479.4	33.6	3133	19 AAX62447	Chlamydia trachoma
7	479.4	33.6	3133	21 AAZ92753	DNA encoding Chlam
8	477.8	33.5	3133	19 AAV40646	Chlamydia trachoma
9	474.6	33.3	3133	7 AAN60007	Sequence encoding
10	397.2	27.9	1038602	20 AAZ01425	Complete genome se
11	352.2	24.7	744	20 AAX25045	Chlamydia psittaci

12 328.8 23.1 1047 21 AAA08124
13 321.8 22.6 726 20 AAX25044
14 287.4 20.2 1362 21 AAA08122
15 287.4 20.2 1371 21 AAA08120
16 287.4 20.2 1542 21 AAA08123
17 254 17.8 1452 21 AAA08121
18 205.2 14.4 831 21 AAA08125
19 171.6 12.0 1909 21 AAX27110
20 140 9.8 720 7 AAN60006
21 140 9.8 720 19 AAV62446
22 140 9.8 720 19 AAV40645
23 140 9.8 720 21 AAZ92752
24 128 9.0 7475 20 AAX02659
25 128 9.0 8192 20 AAX02660
26 125 8.8 7108 20 AAX02662
27 120.4 8.4 4912 20 AAX08423
28 120.4 8.4 4912 21 AAA71402
29 120.4 8.4 5108 22 AAC86258
30 120.4 8.4 5108 22 AAC86259
31 120.4 8.4 5111 22 AAC86262
32 120.4 8.4 5185 22 AAC86254
33 120.4 8.4 5188 22 AAC86266
34 120.4 8.4 5254 22 AAC86267
35 120.4 8.4 5322 21 AAA63225
36 120.4 8.4 5480 20 AAX81562
37 120.4 8.4 5480 20 AAX81562
38 120.4 8.4 7001 20 AAX89795
39 120.4 8.4 7003 20 AAX59390
40 120.4 8.4 7073 20 AAX89796
41 120.4 8.4 7073 20 AAX59391
42 120.4 8.4 7285 20 AAX89797
43 120.4 8.4 7285 20 AAX59392
44 118.8 8.3 7272 20 AAX59393
45 118 8.3 4928 17 AAT32656

ALIGNMENTS

RESULT 1
AAX25047
ID AAX25047 standard; DNA; 1261 BP.
XX
AC AAX25047;
XX
DT 05-JUL-1999 (first entry)
XX
DE Chlamydia psittaci major outer membrane protein DNA.
XX
KW Major outer membrane protein; MOMP; psittacosis; infection;
KW vaccine; genetic immunisation; ss.
XX
OS Chlamydia psittaci.
XX
FH Key Location/Qualifiers
FT CDS 80..1249
FT /*tag= a
XX
PN WO9910005-A1.
XX
PD 04-MAR-1999.
XX
PF 28-AUG-1998; 98WO-US17943.
XX
PR 28-AUG-1997; 97US-0057147.
XX
PA (LOU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.
XX
PI Baghian A, Chouljenko VN, Kousoulas KG, Tully TN;
DR WPI; 1999-254214/21.
DR P-PSDB; AAW98188.
XX

This DNA sequence codes for the major outer membrane protein (MOMP, *AAW98187*) of *Chlamydia psittaci* strain 6BC. Claimed MOMP polypeptides (see *AAW98183* and *AAW98184*) comprise regions VD3 and VD4 of an MOMP, i.e. they lack regions VD1 and VD2. Claimed vaccine compositions include such MOMP polypeptides, optionally fused to a maltose binding protein. Also claimed are isolated nucleic acids encoding the polypeptide, a vector, and a method of preventing C.

CC psittaci infection by administering the vaccine containing the
CC MOMP polypeptide. Vectors encoding MOMP polypeptides lacking
CC regions VD1 and VD2 are useful for genetic vaccination. The
CC vaccines are used to prevent *C. psittaci* infection, especially in
CC birds.

Sequence 1660 BP; 492 A; 342 C; 338 G; 488 T; 0 other;

Query Match 36.9%; Score 526.4; DB 20; Length 1660;
Best Local Similarity 69.5%; Pred. No. 2.7e-153;
Matches 792; Conservative 0; Mismatches 306; Indels 42; Gaps 4;

Qy	129	tgcctgtagggaacccctctgataccaagcttattataatgatgtgatacaatatgggaaggt	188
Db	430	ttgcctgtagggaacccagctgaaccaagttatttaa tcgatggcaactatgtggaaaggt	489
Qy	189	gc tgcagagagactctgcgatactctgcgctacttggtgcgcagctattagtctacgtct	248
Db	490	gtctcaaggagactctgcgatactctgcgctacttggtgcgcagctattagtctacgtct	549
Qy	249	ggattttacggagactatgtttctcacogctattcttaaagttagatgcacotaaacattt	308
Db	550	ggatactacggagattatgtttctgatcgtgtattcaaaagttgatgtaataaaactttt	609
Qy	309	tctatgggagcaagcctactggataccctgctgcacaaactatactact-----	356
Db	610	agcgcgatggctgcgaactcctcagcagcgtacaggtaacgcaagtataactaatcagcca	669
Qy	357	---gccgtagatagacctaaccgcgcctacaataagcatttacacgatgcagagtgttc	413
Db	670	gaagcaaatggcagacgaacatcgcttacggaagccata tgcgaagtgcagagtgttt	729
Qy	414	actaatgcaggcttcattgccttaaacatttgggatacgtcttgatgtttctgtacttta	473
Db	730	tc aaatgcgcgcctctcagccttaaacatttgggatacgtcttcgcacatttctgcacotta	789
Qy	474	ggagctcttaagtgttcacattagagaaactctacagcgtttccaactcgtttggtttatc	533
Db	790	ggggatccaaatggatacttcaagcaagttcgcgctgattccaacttggttgggttaata	849
Qy	534	ggagtttaagagt-----actactgtaaatgcgaatgaactaccacaaagctttct	581
Db	850	gggttttcagctgcgaagctcaactctccagatctccaatgcgaactctcaacgtagc	909
Qy	582	ttaagtaacggagttgtgaactttacacagacacctttctcttgagcgttaggcgt	641
Db	910	attacccaaggtgtgtggaaatttatcacagacacatcttcttgagcgttaggtgca	969
Qy	642	cgtggagccttatgggaa tgcggttgcgaactttgggagcgtgaattccaatatgcacg	701
Db	970	cgtggagccttatgggaa tgcggttgcgaacttttaggagcgtgagttccaatacgc tca	1029
Qy	702	tccaaaccttaagtgaagaaccttaatgtatctgtaaatgcgaatctctgtgaaac	761
Db	1030	tctaa tcttaagatgaattgtcgaatctcaactccaagccagacacattgtgtatccac	1089
Qy	762	aaaccaagggctataaagcgctt-----gc tttcccttgcacacagacgtcgcgta	815
Db	1090	aaaccaagggctataaagggtagctgcgaatttctcttaactataaaggctggaaca	1149
Qy	816	gcaacgctactggaacaaagctcgcgacctcaattatcatgaatggcaagtaggagcc	875
Db	1150	acgaaggtctacagacaccaaa tcaagtcaaatataatcaccatgaatggcaagtaggcctc	1209
Qy	876	tctctatcttacagactaaactctttagtgccaacattttgagtlacaaatggctcgcagca	935
Db	1210	gcctgtcttacagattgaatgtctgttccatata ttgcgtaaactggctcaagaca	1269
Qy	936	acttttgatgctgataaacatccgatactgctcagccaaactacctaagctgtttttaaac	995
Db	1270	acttttgatgctgataactacgtactcgcgtacactaatttcaactaaattaaaatcggaacttcttaac	1329

[illegible]

RESULT 3
AAX25046
ID AAX25046 standard; DNA; 1209 BP.

AA
AC
-
AAX25046;

DT 05-JUL-1999 (first entry)

Chlamydia psittaci major outer membrane protein DNA.

Major outer membrane protein; MOMP; psittacosis; infection;
vaccine; genetic immunisation; ss.

Chlamydia psittaci.

PN WO9910005-A1.

AA
PD
04-MAR-1999.

AA
PF 28-AUG-1998; 98WO-US17943.

PR 28-AUG-1997; 97US-0057147.

PA (LOUU) UNIV LOUISIANA & AGRIC & MECH COLLEGE.

PI Baqhian A, Chouljenko VN, Kousoulas KG, Tully TN;

DR WPI; 1999-254214/21.

XX
F F3DB, HAW30107

XX
FI
FI
A new vaccine for Chlamydia psittaci infections.

XX
10
disclosure, page 22, app, exhibit

and a major outer membrane protein (MOMP) (see AA98187) of Chlamydia psittaci strain LSUWTKC, a cockatiel isolate (the MOMP gene sequence of this isolate is identical to that of C. psittaci Avian Type C). A claimed MOMP polypeptide (see AA98183) comprises regions VD3 and VD4 of MOMP, i.e. it lacks regions VD1 and VD2. A claimed vaccine composition includes MOMP polypeptide lacking VD1 and VD2, optionally fused to a maltose binding protein. Also claimed are an isolated nucleic acid encoding the polypeptide, a vector, and a method of preventing C. psittaci infection by administering the vaccine containing the MOMP polypeptide. Vectors encoding MOMP polypeptides lacking regions VD1 and VD2 are useful for genetic vaccination. The vaccines are used to prevent C. psittaci infection, especially in birds.

Sequence 1209 BP; 352 A; 263 C; 260 G; 334 T; 0 other;

Query Match	Score	DB	Length
Best Local Similarity	36.7%	523.2	20
Matches	69.3%	Pred. No. 2.2e-152	
Matches 790; Conservative	0	Mismatches 308	
		Indels 42	Gaps 4

129	QY	ttgccttaggaaaccccttctgatccaaagcttatttaattgatggtacaataatggaaggt	188
67	Db	ttgccttaggaaacccagctgaaccaagtttatttaatcgatggcactatggaaggt	126
189	QY	gctcagagagatccttgcgatccttgcgtacttgcgtacttgctgcagcctattagcttaagctgt	248
127	Db	gcttcaggagatccttgcatccttgcgtacttgcgtacttgctgcagccatttagcatccgcgcga	186
249	QY	ggattttacggagactatgttttcgacgcttatcttaaaagtagatgcacctaaacattt	308
187	Db	ggatactacggagattatgttttcgatcggtattaaaagttgatgtgaataaaactttt	246
309	QY	tctatggggccaaagcctactggatcgcgtgcgtgcgaacctatactact-----	356
247	Db	agcggcatgggtgcacatctctacgcagctacaggtacaggttaactatactaatcagcca	306
357	QY	---ccgctagatagacctaaacccgcttacaataagcatittacacgatgcagagtgttct	413
307	Db	gaagcaaatggcagaccgaaacatcgcttacgaagagcatatggaaagatgcagagtgtt	366
414	QY	actaatgcagagcttattgcctttaaacatttgggatgcgtttgatgttttctgaacttta	473
367	Db	tcaaatgcagccttctcagctttaaacaatttgggatgcgttcgcactttactcgcacctta	426
474	QY	ggagctttcaatggtttacattagagaaactctacagcttccaactcgtttggtttattc	533
427	Db	ggggcatccaatggatactccaagcaagttcgctgcattccaacttggttgggttaata	486
534	QY	ggagttaaaggt-----actactgtaaatgcaaatgaactacaacaaagctttct	581
487	Db	gggttttcagctgcaagctcaactctcaccatcttccaagcgaactctcctcaacgtaggc	546
582	QY	ttaagttaacggagtgttgaactttacacagacacactttctcttgagcgttaggcgt	641
547	Db	attaccocagggtgtgtggaaatttatacagacacatcatttcttggagcgttaggtgca	606
642	QY	cgtggagccttatgggaatgcggttgtgcgaacttgggagctggaattccaatatgcacag	701
607	Db	cgtggagccttatgggaatgtgtgtgcgaacttaggagctgagttccaatacgcgtcaa	666
702	QY	tccaacacctaaagtgaagaacttaatgctgtatacgtatgcgaattctctgttaaac	761
667	Db	tctaatctctaagattgaaatgctccaagctcaactccaagcccaagcacaatttgtgattcac	726
762	QY	aaacccaaggcctataaaaggcgtt-----gctttcccttgccaacagacgctggccta	815
727	Db	aaaccaagagcctataaaaggagtagctcgaattttctttaccataacgcgtgaaaca	786
816	QY	gcaacagctactggaacaagctcgcacatcaattatcatgaatggcaagttagagcgc	875
787	Db	acagaagctacagacaccaaatcagctcaaatataataccatgaatggcaagttaggcctc	846
876	QY	tctctattctacagactaaactcttagtgccaatacatgtggagtacaaatgggtctcagca	935
847	Db	gcctctgtctacagattgaatafgctgtgtcccatatattggcgtaaacgttgccaagaca	906
936	QY	acttttgatcgtataacatccgcattgtccagccaaaactctcagctgtttttaaac	995
907	Db	acttttgatcgtataccgcatttgcctccaccttaataataaactcgagattcttaac	966
996	QY	ttaactgcattggaacctcttctactaggaaatgccacagcatgtgtctactact-----	1049
967	Db	attactacatggaacccaagccttataggatcaacacactgcttggcccaataatagtggt	1026
1050	QY	---gattcgtctcagacttcaatgcaaatgtttctcgtccagatcaacaagttaaatct	1106
1027	Db	aaggatgtctctatgattcgttcgaattgtctcgtattcagatccaacaaaatgaagtct	1086
1107	QY	agaaaaagct	

Db	1147	actgttgtagcagcgttaatcaatgaagagctgctcacagaatgctcaattcagattc	1206
RESULT	4		
AAA64764			
ID	AAA64764	standard; DNA; 1578 BP.	
XX			
XX	AAA64764;		
XX			
DT	02-FEB-2001	(first entry)	
XX			
DE	C. pneumoniae serovar MOMP5 pmp gene Ral2 fusion coding sequence.		
XX			
KW	Chlamydial infection; sexually transmitted disease;		
KW	pelvic inflammatory disease; PID; tubal obstruction; infertility;		
KW	trachoma; blindness; acute respiratory tract infection;		
KW	Chlamydia; pneumoniae; heart disease; antibacterial; ss.		

RESULTS

AAA64764 standard; DNA; 1578 BP.
 AAA64764;
 02-FEB-2001 (first entry)
 C. pneumoniae serovar MOMPS pmp gene Ral2 fusion coding sequence.
 Chlamydial infection; sexually transmitted disease;
 pelvic inflammatory disease; PID; tubal obstruction; infertility;
 trachoma; blindness; acute respiratory tract infection;
 atherosclerosis; coronary heart disease; antibacterial; ss.
 Chlamydia pneumoniae.
 WO200034483-A2.
 15-JUN-2000.
 08-DEC-1999; 99WO-US29012.
 08-DEC-1998; 98US-0208277.
 08-APR-1999; 99US-0288594.
 01-OCT-1999; 99US-0410568.
 22-OCT-1999; 99US-0426571.
 (CORI-) CORIXA CORP.
 Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
 WPI; 2000-431303/37.
 Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 comprises immunogenic portion of Chlamydia antigen, which comprises
 amino acid sequence encoded by polynucleotide sequence -
 Claim 1; Page 205; 256pp; English.
 The present invention relates to new nucleic acid sequences and the
 proteins encoded by the nucleic acid sequences. The encoded proteins
 comprise an immunogenic portion of a Chlamydia antigen. The encoded
 proteins are useful for the serodiagnosis and treatment of Chlamydia
 infection. Chlamydiae are intracellular bacterial pathogens that are
 responsible for a wide variety of human infections. C. trachomatis
 infection is one of the most common sexually transmitted diseases and can
 lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 and infertility. Trachoma due to ocular infection with C. trachomatis is
 the leading cause of preventable blindness worldwide. C. pneumoniae is a
 major cause of acute respiratory tract infections in humans and is also
 thought to play a role in the pathogenesis of atherosclerosis and
 coronary heart disease. The present sequence is a nucleic acid sequence
 isolated in the present invention.
 Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;

Query Match	36.1%	Score 514.6;	DB 21;	Length 1578;
Best Local Similarity	68.8%	Pred. NO. 1.2e-149;		
Matches 785; Conservative	0;	Mismatches 314;	Indels 42;	Gaps 4;

Qy	128	gttcctctgtagggaacccctctctgatccaaagcttattaaattgatctgggtacaaatatgggaag	187
Db	435	gctgcctctgtagggaacccctctctgatccaaagcttattaaattgatctgggtacaaatatgggaag	494
Qy	188	tgctgcaggagagactcttgcgactcttgcgactcttggcgacgctattattgcttaccgctgc	247
Db	495	tgcttcaggagaactcttgcgactcttgcgactcttggcgacgctattattgcttaccgctgc	554

QY 248 tggattttacgagactgttttcgaccgttatcttaaaagttagatgcacctaataacatt 307
 DB 555 aggatactacgagattattgttttcgactgtattaaagttagatgaataaaacttt 614
 QY 308 ttctatggagccagcactactggttcgctgctgcaactataact- 356
 DB 615 tagcgcatggctgaactctacgcaggtctagtgtaacgaataactaacgacc 674
 QY 357 ----gccgttagatagaccctacccgctcacaataagcatttacagcagatgatt 412
 DB 675 agaagcaaatggcagacgcgaactcgtctacggaaggcattatgcagatgcagatgatt 734
 QY 413 cactaatgcaggtcttcctgcttaaacatttggatgcgttggattttctctgacttt 472
 DB 735 ttcaaatgcagcttcctgacttaaacatttggatgcgttggattttctctgacttt 794
 QY 473 aggaacttctaattgtttacattagaggaaactctacagcgttcaactcgtttgtttatt 532
 DB 795 agggcattccaatggatgatactcaagcaggttgcgctcattcaactgtttgggttaatt 854
 QY 533 cggaggttaaggt-----actactgtaaatgcaaatgcaactaccacacgtttc 580
 DB 855 aggttttcagctgcaagctcaactctctacagatcttccaatgcaacttctcaactagg 914
 QY 581 tttaagttaacgaggttggatttgaactttacacagacacactcttcttctgagcgtagg 640
 DB 915 cattaccacaggtgtgtggaattttatatacagacacatcattttcttggagcgtagg 974
 QY 641 tctgagaccttatgggaatgcggtgtgcaacttggagcgtgaattcccaatatgcaca 700
 DB 975 acgtggagcttctatgggaatgtgtgtgcaacttggagcgtgaattcccaatatgcaca 1034
 QY 701 gtccaaacttaagtgaagacttaattgtatctgtaacgtatcgcaattctctgtataa 760
 DB 1035 attcaacttaagttagatgctcaacgtcaacttcaagccagacacaaatttggatca 1094
 QY 761 caaacccaaaggtctataagcgtt-----gcttcccttgcacacagcgtggtggt 814
 DB 1095 caaaccaagaggtctataagcgttagctcgaaatttcttctacataaagcgtggaac 1154
 QY 815 agcaacagctactggaacaaagctgcgaccatcaattatcatgaatggcgaagtaggagc 874
 DB 1155 aacagaagctacagacacacaaatcagctacataataacacatgaatggcgaagtagg 1214
 QY 875 ctctctatctacagactaaactcttttagtgcatacatttggatgacatggtctcgagc 934
 DB 1215 cgccctgtctacagatgaatgtgtgttccatattatggtgcaactggtcaagagc 1274
 QY 935 aacttttgatgctgatacactcgcgcatgtctcagccaaactacacacagctgtttataa 994
 DB 1275 aacttttgatgctgatacactcgcgcatgtctcagccaaactacacacagctgtttataa 1334
 QY 995 cttaactgcatggaaccccttcttactaggaatgcccacagctgttctactact- 1049
 DB 1335 cattactacatggaaccccaagccttataggatcaaccactgctttgcccataatagtc 1394
 QY 1050 ----gattcctctcagactctcaaatgttttctgctcagatcaacaagtttataatc 1105
 DB 1395 taagatgttctatctatgtctgtcgaattgtctcagatcagatcaacaagaatgagtc 1454
 QY 1106 tagaagaactgtgagttactgtatgagcactcttattgtatgctgataaagtgcact 1165
 DB 1455 tagaagaactgtgagttactgtatgagcactcttattgtatgctgataaagtgcact 1514
 QY 1166 tactcagaagctgttttaataacagagagagcgtcgtcactgtatgttcagttcagatt 1225
 DB 1515 cactggtgaagcagcgttaatacaatgaagagcgtcgtcactgtatgttcagttcagatt 1275
 QY 1226 c 1226
 DB 1575 c 1575

RESULT 5
 AAH56267
 ID AAH56267 standard; DNA; 1578 BP.
 XX
 AC AAH56267;
 DT 05-SEP-2001 (first entry)
 XX
 DE Chlamydia trachomatis pmp gene sequence.
 XX
 KW Chlamydia; vaccine; infection; fusion protein; antigen; heart disease;
 KW pelvic inflammatory disease; trachoma; atherosclerosis; heart disease;
 KW acute respiratory tract infection; Cap1; Ctf529; OMCB;
 KW polymorphic membrane protein; pmp; thiol specific antioxidant; TSA; ds.
 XX
 OS Chlamydia trachomatis.
 XX
 WO200140474-A2.
 XX
 PD 07-JUN-2001.
 XX
 PF 04-DEC-2000; 2000WO-US32919.
 XX
 PR 03-DEC-1999; 99US-0454684.
 PR 19-APR-2000; 2000US-0556877.
 PR 20-JUN-2000; 2000US-0598419.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Probst P, Bhatia A, Skeiky YAW, Fling SP, Scholler J;
 XX
 WIPI; 2001-374831/39.
 XX
 PT Chlamydia polypeptides and fusion proteins useful for preventing pelvic
 PT inflammatory disease, trachoma, acute respiratory tract infections,
 PT atherosclerosis and heart disease -
 XX
 PS Claim 1; Page 211; 295pp; English.
 XX
 CC The present nucleotide sequence is provided in a specification
 CC relating to compounds and methods for the treatment and diagnosis of
 CC chlamydial infection. The compounds provided include polypeptides and
 CC fusion proteins comprising immunogenic portions of Chlamydia antigens
 CC and DNA sequences encoding such polypeptides. They are useful for
 CC vaccinating against chlamydial infection, which causes pelvic
 CC inflammatory disease, trachoma, acute respiratory tract infections,
 CC atherosclerosis and heart disease.
 XX
 SQ Sequence 1578 BP; 414 A; 394 C; 382 G; 388 T; 0 other;

Query Match 36.1%; Score 514.6; DB 22; Length 1578;
 Best Local Similarity 68.8%; Pred No. 1.2e-149;
 Matches 785; Conservative 0; Mismatches 314; Indels 42; Gaps 4;
 QY 128 gttgctgtgaggaaccccttctgacccaagcttattatgattgataatggaag 187
 DB 435 gctgctgtgggaatccagctgaacaaatttattatgctgacactatgtgggaag 494
 QY 188 tgcctgaggaactcttgcgactcttgcgacttgcgacgctattgcttacgtgc 247
 DB 495 tgcctgaggaactcttgcgactcttgcgacttgcgacgctattgcttacgtgc 554
 QY 248 tggattttacgagactatgttttcgaccgtatctttaaagtagatgcacctaataact 307
 DB 555 aggatactacgagattattgttttcgactgctgattataaagattgattgaataaaacttt 614
 QY 308 ttctatggagccagcactactggtatcgctgctgcaactataactact- 356
 DB 615 tagcgcatggctgaactctacgcaggtctagtgtaacgaataactaacgacc 674
 QY 357 ----gccgttagatagaccctacccgctcacaataagcatttacagcagatgatt 412

Db 675 agaagcaaatggcagaccgaacatcgtcttaacggaagcatatgcaagatgcagatgggt 734
Qy 413 cactaaatgcagcttcattgcttaaacatttggatgcgttttctgtacttt 472
Db 735 ttcaaatgcagcttcattgcttaaacatttggatgcgttttctgtacttt 794
Qy 473 agggacttctaatgtttacattagagaaactctacagcgttcaatctgtttgtttatt 532
Db 795 agggatcccaatgatacttcaaacgaagtgcgttcgcattoactgttgggttaatt 854
Qy 533 cggagcttaaggtt-----actactgtaaatgcaaatgaactacaaacgtttc 580
Db 855 aggggtttcagctgcaagctcaatctctacagatcttccaaatgcaacttctcaacgtagg 914
Qy 581 tttaagtaacggaggtgtgtgaactttacacagacacgtcttctctgtggagcgttaggcgc 640
Db 915 cattaccgaaggtgtgtggaattttatacagacacatcattttctggagcgttaggtgc 974
Qy 641 tctgtgagccttatgggaatggtgtgtgcaactttggagcctttggagcgttaggcgc 700
Db 975 acgtggagctttatgggaatggtgtgtgcaacttttagagcgttaggttccaaatgcgtca 1034
Qy 701 gtccaaacttaagtgtgaagaacttaagtgtgtgttaacgtatgcgaattctctgttaa 760
Db 1035 atctaactgaagtgtgagctgcaacgtcacttcaagccagcagacaaatttggattca 1094
Qy 761 caaacccagggtctaaagggtt-----gctttcccttgcacacagcgtggcgt 814
Db 1095 caaaccaagggtctaaagggtctgctgaatttcttaccataaagcgtgggaac 1154
Qy 815 agcaacagctactgcaacaaagctgtgacacatcaattatcatgaatggcgaagtaggagc 874
Db 1155 aacagaagctacagacacaaatcagctacaaattaaatcaccatgaatggcgaagtaggct 1214
Qy 875 cctctatcttacagactaaacttttagtgcatacattggagtcacaaatgtctcgagc 934
Db 1215 cgcctctcttacagatgaatatctgtgtccatatattggcgtaaactgttcaagagc 1274
Qy 935 aactttgtatgctataacatccgcatgtgtcagcacaactcactcagctgtttttaa 994
Db 1275 aactttgtatgctatacctcgcattgtcctcaactaaattaaatcggagatcttaa 1334
Qy 995 cttaactgcatgaaccccttttactaggaatgccacagcattgtctactact- ---- 1049
Db 1335 cattactatggaaaccccaagccttataggatacaaccacgtgttggcccaataatagtgg 1394
Qy 1050 ----gattcgttcagacttcaatgcaaatgtttctgtcagatcaacaaagttaaactc 1105
Db 1395 taaggatgtctatcgtatgtcttgcaaatgtcttcagatcaacaaatgaagtc 1454
Qy 1106 tagaaagctgtgtgagttactgttagagactacttttagtgcgtgataaatgtcact 1165
Db 1455 tagaaagctgtgtgagttactgttagagactacttttagtgcgtgataaatgtcact 1514
Qy 1166 tactcagagcgtgtttaaataacagagagcgtcgcacgtatctgttcagttcagatt 1225
Db 1515 cactgttgagacagcgtttaaataacagagagcgtcgcacgtatctgttcagttcagatt 1574
Qy 1226 c 1226
Db 1575 c 1575

XX KW Major outer membrane protein; MOMP; diagnosis; vaccine; ds.
XX OS Chlamydia trachomatis serovar L2.
XX FH Key Location/Qualifiers
FT CDS 1288..2472
FT sig_peptide /*tag= a
FT mat_peptide /*tag= b
FT /*tag= c
XX US5821055-A.
XX 13-OCT-1998.
XX 06-JUN-1995; 95US-0468451.
XX 13-JAN-1986; 86US-0818523.
XX 14-JAN-1985; 85US-0692001.
XX 25-APR-1991; 91US-0691639.
XX 28-OCT-1993; 93US-0144095.
XX 06-JUN-1995; 95US-0468451.
XX (CHIR) CHIRON CORP.
XX (WASH-) WASHINGTON RES FOUND.
XX Agabian N, Kuo C, Mullenbach G, Stephens R;
XX WPI; 1998-567652/48.
XX P-PSDB; AAW73141.
XX Probe for detecting Chlamydia trachomatis - comprises polynucleotide
XX fragment that hybridises to major outer membrane protein DNA or RNA
XX Claim 5; Fig 2A-E; 15pp; English.
XX This DNA sequence codes for a 42 kDa major outer membrane protein
XX (MOMP, see AAW73141) of Chlamydia trachomatis serovar L2. A library
XX of chlamydial genomic DNA was produced in the phage lambda 1059
XX system. A lambda 1059 recombinant having a 9.2 kb insert was shown
XX to be homologous to lambda gt10/33 (see AAV62446) by Southern
XX analysis and was used for endonuclease digestion mapping and
XX additional Southern analyses. 2 Contiguous fragments were
XX identified and these contained sufficient base pairs to encode the
XX L2 MOMP gene product. These fragments were cloned into M13 for
XX DNA sequencing. Novel recombinant DNA constructs are provided for
XX the expression of a polypeptide having immunological activity
XX corresponding to that of a naturally-occurring MOMP of C.
XX trachomatis. Such polypeptides find use as reagents in the
XX detection of C. trachomatis, or antibodies to C. trachomatis, and
XX as vaccines against infection by C. trachomatis in susceptible
XX hosts. A claimed probe for detecting C. trachomatis comprises a
XX polynucleotide fragment that specifically hybridises to a DNA or RNA
XX sequence encoding C. trachomatis 38-45 kDa MOMP. The probe has
XX a sequence which is complementary to at least 12 contiguous bases
XX of the sequence given in AAV62447.
XX Sequence 3133 BP; 911 A; 667 C; 611 G; 944 T; 0 other;

Query Match 33.6%; Score 479.4; DB 19; Length 3133;
Best Local Similarity 66.5%; Pred. No. 1.6e-138;
Matches 741; Conservative 0; Mismatches 356; Indels 18; Gaps 3;
Qy 130 tgcctgtagggaacccttctcaatcgaagcttattgagtggttcaatattgggaaggtg 189
Db 1355 tgcctgtagggaacccttctcaatcgaagcttattgagtggttcaatattgggaaggtg 1414
Qy 190 ctgcaggagatccttgcgactccttgcgacttctgtgcgagcgtattgcttactgtcgtg 249
Db 1415 tcggcggagatccttgcgactccttgcgacttctgtgcgagcgtattgcttactgtcgtg 1474

RESULT 6
AAV62447
ID AAV62447 standard; DNA; 3133 BP.
XX AC
XX AC AAV62447;
XX AC
XX DT 02-FEB-1999 (first entry)
XX DE Chlamydia trachomatis major outer membrane protein DNA.

```
QY 250 gattttacgagactatgttttcgaccgtatctttaaagtagatgcacctaataacattt 309
D 1475 gtactatggtgactttgttttcgaccgtgttttgaacagatgtgaataagaattcc 1534
QY 310 ctatgggagccaagccta-----ctggtccgctgtgcaactatactactgcgtag 363
D 1535 aaatgggtgccaagcctaactgctacaggaatgctgcagctccatccactgtacag 1594
QY 364 atagacctaacccgctcacaataagcatttacagatgcagagtggttactaatatgag 423
D 1595 caagagaaatcctgtctacgcccacatagcagatgctgagatgtttacaaatgtg 1654
QY 424 gcttcgtcttaaacatttggtgctctttgtgatttttctgtactttaggagcttcta 483
D 1655 ctacatggcattgaatttggtgctgttttgatgtattctgtacattaggagccacca 1714
QY 484 atggttacattagagaaactctacagcgttcaatctggttgggtttatcgaggtta--- 540
D 1715 gtggatatcttaaggaattcagctcttccaactgtttggttattcgagagataatg 1774
QY 541 -----aaggtactactgtaaatgcaaatgaactaccacaaagctttcttaagtaacgag 594
D 1775 aqaacactgtcagatttcagatagtaagctgtaccacaaatagagcttagatcaattg 1834
QY 595 ttgttgaactttacacagacacctcttcttggagcgtgagcgtcgtgagccttat 654
D 1835 ttgttgagttgtatacagatactacttttctgtgagtgctgagcgtcgtgagccttat 1894
QY 655 ggaagtcggtgtgcaactttgagcgtgaattccaatagcagacagtcacacacctaag 714
D 1895 gggaaatggtgagcgtcttagcgtcttcttccaatacagctcaatcccaagcctaag 1954
QY 715 ttgaagaaacttaagtgtatgtatgacgtatgcgaattctctgtaacaaacccaagcct 774
D 1955 tgaagaattaaacgttctctgtaacgagcgtgatttactatcaataagccttaagat 2014
QY 775 ataaaggcgttcttcccttgcacacagcgtggtggtgagcagcactactggaacaa 834
D 2015 atgtagggaagaattcccttgcatttaaaagcaggaacagatggtgtgacaggaacta 2074
QY 835 agtctgcgacatcatatcatgaatgagcaagtaggagcctctctctctacagactaa 894
D 2075 aggatgctctattgattaccatgaatggaagcaagtttagctctctctctacagactga 2134
QY 895 actctttagtgccatcacattgagtagcaatggtctcgagcaacttttgcgtgataaca 954
D 2135 atatttcaactccctacattgagttaaatggtctcgagcaagtttgcgtgacagacga 2194
QY 955 tccgattgctcagcgaacaaactacacagcgtgttttaacttaactgagcagacccctt 1014
D 2195 ttctgattgctcagcgaagtcagctacacagctgtcttgcgtgttaccactctgaccccaa 2254
QY 1015 ctctactaggaatgccacagcattgtctactactga---ttcgttctcagactctcagc 1071
D 2255 ctatgctgagcgtggcagtgaaagctagcgcagaggggtcagctcgagataccatgc 2314
QY 1072 aaattgttctctgcagatcaacagtttaatactagaaaagcgttgcgtgactgtag 1131
D 2315 aaatggttctctgcgaattgaacagatgaatactagaaaactctgcgtgattgcagtag 2374
QY 1132 gagctactttagtgcataaattggtcacttactgcagaaagcgttgaataaag 1191
D 2375 gaacaactattgtgagtcagacaaatgcagattgacagttgagactcgttcgtagc 2434
QY 1192 agagagcgtcagctatctgtgctcagttcagattc 1226
D 2435 agagagcgtcagctaaatgcacattccgcttc 2469
```

RESULT 7
AAZ92753
ID AAZ92753 standard; DNA; 3133 BP.

```
XX AAZ92753;  
XX AC  
XX 19-JUN-2000 (first entry)  
XX DNA encoding Chlamydia trachomatis major outer membrane protein (MOMP).  
XX Major outer membrane protein; MOMP; serovar L2; immunoreactive;  
XX antibody production; immunoassay; detection; vaccine; trachoma;  
XX inclusion conjunctivitis; pneumonia; lymphogranuloma venereum;  
XX mucous membrane genital tract infections; ds.  
XX Chlamydia trachomatis serovar L2.  
XX Key Location/Qualifiers  
XX CDS 1288..2472  
XX /*tag= a  
XX /product= "Chlamydia trachomatis serovar L2 MOMP"  
XX US6030799-A.  
XX 29-FEB-2000.  
XX 06-JUN-1995; 95US-0466152.  
XX 28-OCT-1993; 93US-0144095.  
XX 13-JAN-1986; 86US-0818523.  
XX 25-APR-1991; 91US-0691639.  
XX 14-JAN-1985; 85US-0692001.  
XX (WASH-) WASHINGTON RES FOUND.  
XX Agabian N, Stephens R, Kuo C-C, Mullenbach G;  
XX WPT; 2000-223163/19.  
XX P-PSDB; AAY81266.  
XX New synthetic or recombinant polypeptide, useful for diagnosing or  
XX preventing Chlamydia trachomatis infection, is immunologically  
XX equivalent to a major outer membrane protein -  
XX Examples; Fig. 2; 17pp; English.  
XX The present sequence represents DNA encoding the major outer membrane  
XX protein (MOMP) of Chlamydia trachomatis serovar L2. DNA obtained from  
XX C. trachomatis serovar L2 was partially digested with Dnase I and  
XX inserted into the bacteriophage vector lambda-gt11; a lambda-gt11 insert  
XX (AAZ92752; ATCC #40157, referred to as lambda-gt11/L2/33 in the  
XX specification) was obtained which encodes a protein fragment (AAY81267)  
XX reactive with a pool of C. trachomatis-specific monoclonal antibodies.  
XX The lambda-gt11/L2/33 insert sequence was also used to probe a  
XX Chlamydia genomic DNA phage lambda 1059 library, and the present  
XX sequence which encodes the full-length Chlamydia trachomatis serovar L2  
XX MOMP was identified. C. trachomatis is a human pathogen responsible for  
XX diseases such as trachoma, inclusion conjunctivitis, pneumonia,  
XX lymphogranuloma venereum, and mucous membrane genital tract infections  
XX such as cervicitis and urethritis. The full-length recombinant MOMP or a  
XX fragment thereof is used to elicit the production of antibodies to a MOMP  
XX of Chlamydia trachomatis. MOMP proteins are also useful as immunoassay  
XX reagents (substitutes for native MOMP) for detecting C. trachomatis or  
XX its antibodies, for diagnosing infection, or as an immunogen for  
XX vaccines.  
XX Sequence 3133 BP; 911 A; 670 C; 609 G; 943 T; 0 other;
```

Query Match 33.6%; Score 479.4; DB 21; Length 3133;
Best Local Similarity 66.5%; Pred. No. 1.6e-138;
Matches 741; Conservative 0; Mismatches 356; Indels 18; Gaps 3;

QY 130 tgcctgtagggaaccccttctgatccaaagcttatttgatggtgacaaatgggaaggtt 189
D 1355 tgcctgtagggaaccccttctgatccaaagcttatttgatggtgacaaatgggaaggtt 1414

XX	AAV40646	standard; DNA; 3133 BP.
XX	AAV40646;	
XX	27-OCT-1998	(first entry)
XX	Chlamydia trachomatis	serovar L2 MOMP coding region.
XX	MOMP;	major outer membrane protein; immunoassay; diagnosis;
XX	detection;	antibody; serovar L2; ds.
XX	Chlamydia trachomatis.	
XX	Key	Location/Qualifiers
XX	CDS	1288..2472
XX	FT	/*tag= a
XX	FT	/product= MOMP
XX	FT	sig_peptide
XX	FT	1287..1353
XX	FT	/*tag= b
XX	US5770714-A.	
XX	23-JUN-1998.	
XX	14-JAN-1985;	85US-0692001.
XX	13-JAN-1986;	86US-0818523.
XX	14-JAN-1985;	85US-0692001.
XX	25-APR-1991;	91US-0691639.
XX	28-OCT-1993;	93US-0144095.
XX	06-JUN-1995;	95US-0466814.
XX	(CHIR)	CHIRON CORP.
XX	(WASH-)	WASHINGTON RES FOUND.
XX	Agabian N,	Kuo C, Mullenbach G, Stephens R;
XX	WPI;	1998-376887/32.
XX	P-PSDB;	AAW57775.
XX	DNA coding for	Chlamydia trachomatis polypeptide - useful for
XX	producing	recombinant polypeptide, etc.
XX	Claim 5;	Fig 2; 15pp; English.
XX	The sequence is	that encoding a major outer membrane
XX	protein (MOMP)	of Chlamydia trachomatis. This polypeptide can
XX	be used in	immunoassays, e.g. to detect Chlamydia trachomatis
XX	antibodies in	blood, or can be used in vaccines. The
XX	polynucleotide	can be labelled and used as a diagnostic probe.
XX	Sequence	3133 BP; 909 A; 670 C; 610 G; 944 T; 0 other;
XX	Query Match	33.5%; Score 477.8; DB 19; Length 3133;
XX	Best Local Similarity	56.4%; Pred. No. 5e-138;
XX	Matches 740;	Conservative 0; Mismatches 357; Indels 18; Gaps
QY	130	tgccgtgtagggaaacccctctgtaccgaagcttatttaattgatgggtacaatatgggaaggtg 189
Db	1355	tgccgtgtagggaaacccctctgtaccgaagcttatttaattgatgggtacaatatgggaaggtg 1414
QY	190	ctgcagagagatcccttgcgatcccttgccgctacitgtgtgcagcattattagcttaccgtgctg 249
Db	1415	ctgcagagagatcccttgcgatcccttgccgctacitgtgtgcagcattattagcttaccgtgctg 1474
QY	250	gatttttcggsagactatgttttcgcgcgtatctctaaagtagatgcacctaaacatttt 309
Db	1475	gttactatgtgacttgttttcgcgcgtatctctaaagtagatgcacctaaacatttt 1534

XX	WO9928475-A2..	
XX		
XX	10-JUN-1999.	
XX		
XX	27-NOV-1998; 98WO-IB01939.	
XX		
XX	04-NOV-1998; 98US-0107077.	
XX	28-NOV-1997; 97FR-0015041.	
XX	17-DEC-1997; 97FR-0016034.	
XX		
XX	(GEST) GENSET.	
XX		
XX	Griffals R;	
XX		
XX	WPI; 1999-371125/31.	
XX		
XX	Genome sequence of Chlamydia trachomatis	
XX		
XX	Claim 1; Page 373-656; 1755pp; English.	
XX		
XX	The present sequence represents the complete genome of Chlamydia	
XX	trachomatis. Open reading frames (ORFs) of the genome encode	
XX	polypeptides AAY36754-Y37949. The polypeptides can be used as vaccines	
XX	against Chlamydia trachomatis. Antisense and ribozyme sequences can also	
XX	be used to control growth of the microorganism. Chlamydia trachomatis is	
XX	responsible for a large number of diseases, e.g. eye diseases such as	
XX	conventional trachoma, nonendemic trachoma, paratrachoma, and inclusion	
XX	conjunctivitis; genital diseases such as nongonococcal urethritis,	
XX	epidymitis, cervicitis, salpingitis, perihepatitis, Bartholinitis;	
XX	enanthopathy in breast feeding infants; and venereal	
XX	lymphogranulomatosis. The polypeptides of the invention may be of use in	
XX	treating these diseases.	
XX		
XX	sequence 1038602 BP:304265 A; 214645 C; 214259 G; 305001 T; 432 other;	
XX		

364	QY	atagacctaacccgctctacaataagcatttacacgatgcagagtgttccactaatcag	423
Db	733547	caagagagaaatcctctcttacggccgaacatatgcaggatgctgagatgtttacaaatgctg	733606
424	QY	gttcattgccttaaacatttgggacgctttgatgtttctgtacttttaggagccttcta	483
Db	733607	cttacatggcattgaattttgggctgttttgatgtattctgtacatttagagccacca	733666
484	QY	atggttacatttagagaaactctacagcgttcaactctgttgggtttattcggagatgaaa	540
Db	733667	gtggatattcttaaggaattacagcatcttcaactttagtctgggtattctcggagataatg	733726
541	QY	-----aaggttactctgtaaatgcaaatgaactaccacaaacgcttctttaaataacggag	594
Db	733727	agaacatgctacagtttcagatagtaagctgtgtaccacaaatagagcttagatcaatctg	733786
595	QY	tgttgaaacttttacacagacaccttttctcttgagcgtgagcgctcgtggaacatttat	654

Db 616 tggtaaatcactgtgtgaagcagcgttaataatcaatgaagagccgctcaatgaatgtctcaa 675

Qy 1218 ttcagattc 1226

Db 676 ttcagattc 684

RESULT 12

AAA08124

ID AAA08124 standard; DNA; 1047 BP.

XX

AC AAA08124;

XX 27-JUN-2000 (first entry)

XX

XX C. trachomatis MOMP containing fusion protein nucleotide sequence #5.

XX Chlamydia trachomatis; fusion protein; major outer membrane protein;

XX MOMP; hydrophilic polypeptide; antibody; detection; diagnosis;

XX infection; infectious disease; ds.

XX

XX Chlamydia trachomatis.

OS Synthetic.

XX

XX Key Location/Qualifiers

XX 1..1047

XX /tag=

XX /note= "Fusion protein containing at least part of a major outer membrane protein (MOMP) of Chlamydia trachomatis; no stop codon given"

FT

FT JP2000041678-A.

XX

PN 15-FEB-2000.

XX

XX 28-JUL-1998; 98JP-0213212.

XX

PR 28-JUL-1998; 98JP-0213212.

XX

XX (ELED) DENKI KAGAKU KOGYO KK.

XX

DR WPI; 2000-295780/26.

DR

DR P-PSDB; AAY82392.

XX

XX A soluble fused protein useful for diagnosis of Chlamydia infection,

XX comprises at least part of major outer membrane protein (MOMP) of

XX Chlamydia trachomatis -

XX Claim 20; Page 29-30; 37pp; Japanese.

XX The present invention describes fusion proteins (I) comprising at least part of a major outer membrane protein (MOMP) of Chlamydia trachomatis, at least one hydrophilic polypeptide having no immunoreactivity to human serum and their connected part. AAA08120 to AAA08125 encode specifically claimed examples of the fusion proteins given in AAY82388 to AAY82393. Also described is a method (A) for the detection of Chlamydia trachomatis antibody using (I) as the antigen. (I) is used for the diagnosis of Chlamydia trachomatis infectious diseases. The method can diagnose Chlamydia trachomatis infectious diseases specifically in a high sensitivity.

XX Sequence 1047 BP; 315 A; 209 C; 233 G; 290 T; 0 other;

XX

XX Query Match 23.1%; Score 328.8; DB 21; Length 1047;

XX Best Local Similarity 60.5%; Pred. No. 6.5e-92;

XX Matches 671; Conservative 0; Mismatches 367; Indels 72; Gaps 5;

Qy 126 atgttgctgtagggaaccctctgatccaagcttattatgtatgtgataatgggaa 185

Db

1 atgtgctgtggtaaccctctgtaaccaagcttatgatgcagcggtatcctatggaa 60

Qy

186 ggtgctgcaggagatccttcgataccttgcgtacttggcgagcgtattagcttacct 245

Db 61 ggttcggcgagatccttgcgactctgcaccacttggtgagcgtatcagcatgct 120

Qy 246 gctggattttaagcgagactatgttttcgaccgtatcttaaaagtagatgcaactaaaca 305

Db 121 atgggttactatgtgacttggtttttcgaccgtgttttgcacacagatgtaataugaa 180

Qy 306 ttttctatgggagccagccta-----ctggatccgctgctgcaaacatactactgccc 359

Db 181 ttcacaaatgggtgccaagcctcaactgctacaggcaatgctgagcgtccatccacttgc 240

Qy 360 gtagatagacctaaccgcgctcacaataagcatttacacgagtcgagagtggttccaat 419

Db 241 acagcaagagagaatcctgttaccgcccacatgacagatgctgagatgtttacaaat 300

Qy 420 gcaggcttcatttccttaaacatttgggacgttggatcgtttgattttctgtacttttagagct 479

Db 301 gctgctacatgcatgtaatttgggacgtttgattttctgtacttttagagagcc 360

Qy 480 tctaaatggttacattagaggaactctacagcgttccaatctggtttatttcgaggtt 539

Db 361 accagtgatattcttaaaaggaatttcagcatcttccaacttagttggttatttcggagat 420

Qy 540 aaaggttactactgtaaatgcaatgaactaccacaaacgtttctttaaagtaacggagttgt 599

Db 421 aatg-----agaacctgctacagtttcagatagtagaagcttgacca 462

Qy 600 gaactttacagacacactcttcttcttggagcgtaggcgtctggtggagccttatggaa 659

Db 463 aatatgagcttagatcaatct----- 483

Qy 660 tgcggttgccaactttgggagctgaattccaatatgacagtcaccaaaccttaagttgaa 719

Db 484 ----gtgttgagttgtatgtaacaataacaataacaataacaataacaataa 539

Qy 720 gaacttaattgtatgtataacgtatcgcaattctcttaaaacccaagggcgtataaa 779

Db 540 caa--taacaataacaataacaatgctgagtttactatcaataagccttaagagatatgta 597

Qy 780 ggcgttcttcccttggccaacagacgtggcgtagcaacagctactggaacaaagtct 839

Db 598 gggcaagaattccctcttgatcttaaaagcaggaacagatggtgacaggaactaagat 657

Qy 840 gcgaccatcaattatcatgaatggcaagtaggagcctctctcttctacagactaaactct 899

Db 658 gcctctattgattaccatgaatggcaagcaggttagctctctcttctacagactgaatg 717

Qy 900 ttagtgccatacattggagtaacaatggtctcgagcaactttttagctgataacacatccgc 959

Db 718 ttcactccctacattggagttaaatggtctcgagcaagttttgtagcagacagatctgt 777

Qy 960 attgctcagccaaactacacagctgttttaacttaactgaactgcatggaacccctcttta 1019

Db 778 attgctcagccagagtcagctacacactgctttgattgttaccactctgaacccaactatt 837

Qy 1020 ctaggaaatgccacagattgtctactactga---ttcgttctcagactctcagactc 1076

Db 838 gctggagctgcgcatgtgaagcagcagcagagggctcgcgtcggagataccatcacaatc 897

Qy 1077 gtttctgtcagatacaacaagttttaaatctagaaaaagcttggagttactgtaggagct 1136

Db 898 gtttctgtcaattgaaacaagatgaaatctagaaaaatctgcggtattgcagtaggaaca 957

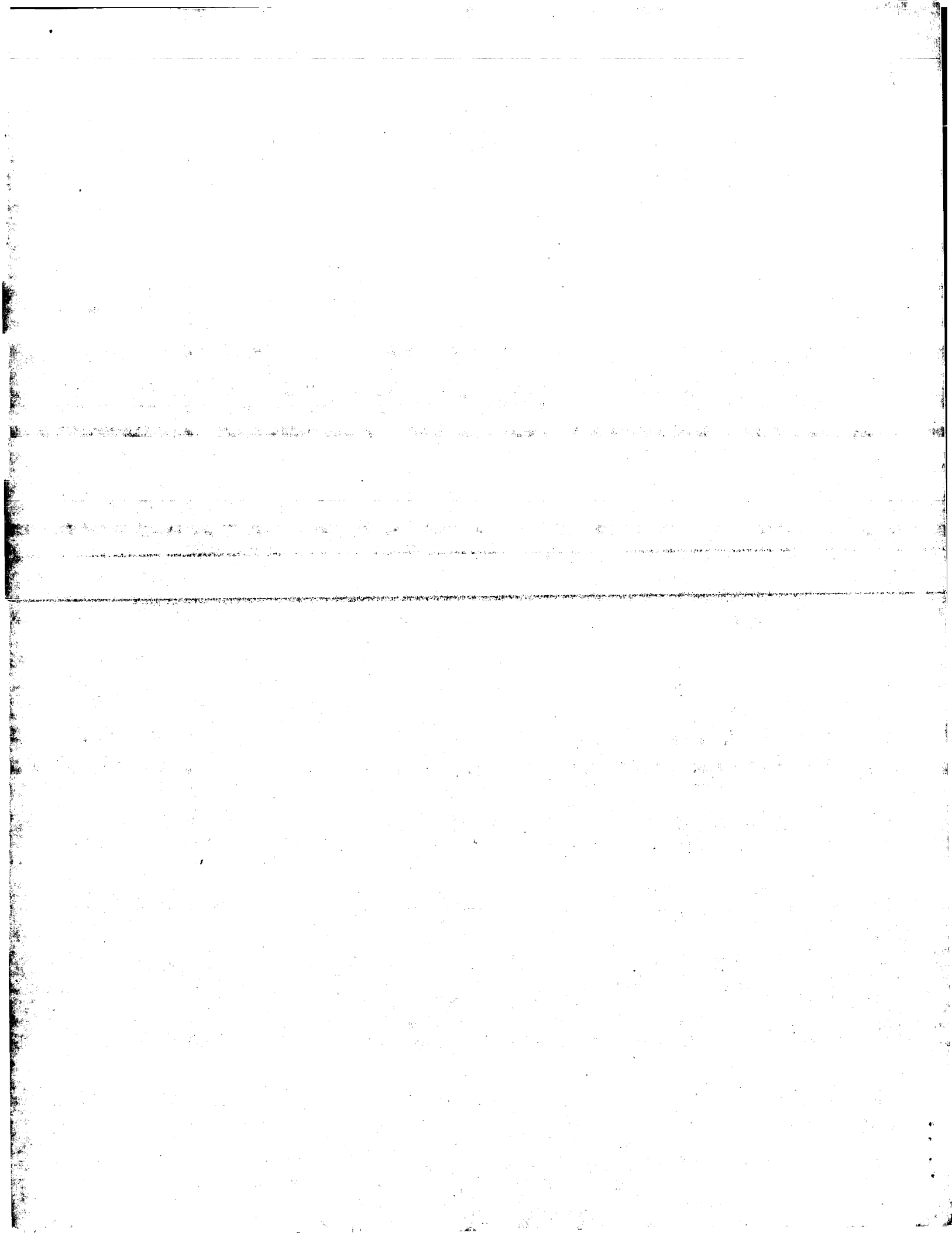
Qy 1137 accttagttgactgtataaatggtcacttactgcagaagcgttttaatttaacagagaga 1196

Db 958 actattgtggatgcagacaaatcacagttacagttgagactcgtctgtagcagagaga 1017

Qy 1197 gctgtcagctatctggtcagttcagattc 1226

Db 1018 gctgtcagcttaaatgcacaattccgcttc 1047

RESULT 13



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:54:07 ; Search time 243.49 Seconds
(without alignments)
1326.369 Million cell updates/sec

Title: US-09-391-606-12

Perfect score: 1426

Sequence: 1 tggacgactctgtgtgc.....tcctgacctggaagtgc 1426

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	118	8.3	4928	1 US-08-345-913-1	Sequence 1, Appl
2	118	8.3	4928	3 US-08-818-562-1	Sequence 1, Appl
3	104.8	7.3	4328	4 US-09-132-808-1	Sequence 1, Appl
4	104.8	7.3	4328	4 US-08-910-647-2	Sequence 2, Appl
5	104.8	7.3	4818	4 US-08-910-647-2	Sequence 4, Appl
6	104.8	7.3	5107	4 US-08-910-647-3	Sequence 3, Appl
7	104.8	7.3	9600	4 US-08-910-647-1	Sequence 1, Appl
8	104.2	7.3	3125	2 US-08-037-816A-13	Sequence 13, Appl
9	103.6	7.3	3125	2 US-08-530-146-13	Sequence 13, Appl
10	103.6	7.3	5676	2 US-08-663-998-3	Sequence 3, Appl
11	103.6	7.3	5682	2 US-08-663-998-4	Sequence 4, Appl
12	103.6	7.3	5900	2 US-08-663-998-1	Sequence 1, Appl
13	103.6	7.3	5952	2 US-08-663-998-2	Sequence 2, Appl
c 14	103	7.2	2057	1 US-08-450-945-57	Sequence 57, Appl
c 15	103	7.2	2057	4 US-08-976-161-57	Sequence 57, Appl
c 16	102.8	7.2	13254	1 US-08-276-852-156	Sequence 156, App
c 17	102.8	7.2	13254	1 US-08-276-852-170	Sequence 170, App
c 18	102.8	7.2	13254	1 US-08-899-575-156	Sequence 156, App
c 19	102.8	7.2	13254	1 US-08-899-575-170	Sequence 170, App
c 20	102.8	7.2	13254	1 US-08-899-575-156	Sequence 156, App
c 21	102.8	7.2	13254	1 US-08-899-575-170	Sequence 170, App
c 22	102.8	7.2	13254	5 PCT-US95-08743-156	Sequence 156, App
c 23	102.8	7.2	13254	5 PCT-US95-08743-170	Sequence 170, App
c 24	102.6	7.2	420	4 US-09-158-863C-64	Sequence 64, Appl
25	97	6.8	4326	4 US-08-760-615-7	Sequence 7, Appl
26	89.8	6.3	6253	2 US-08-627-151A-5	Sequence 5, Appl
27	86	6.0	3853	3 US-08-801-092-5	Sequence 5, Appl

28	86	6.0	4026	3 US-08-801-092-19	Sequence 19, Appl
29	86	6.0	4249	3 US-08-801-092-33	Sequence 33, Appl
30	86	6.0	4283	1 US-08-343-401A-3	Sequence 3, Appl
31	86	6.0	4283	1 US-08-445-265A-1	Sequence 1, Appl
32	86	6.0	4283	3 US-08-990-442-1	Sequence 1, Appl
33	86	6.0	4965	2 US-08-564-313-1	Sequence 1, Appl
34	86	6.0	4965	5 PCT-US94-06069-1	Sequence 1, Appl
35	86	6.0	5653	1 US-08-073-836-3	Sequence 3, Appl
36	86	6.0	5653	1 US-08-235-277-1	Sequence 1, Appl
37	86	6.0	6238	3 US-08-646-538-5	Sequence 5, Appl
38	86	6.0	6238	4 US-09-503-222-5	Sequence 5, Appl
39	86	6.0	6253	3 US-08-893-327-15	Sequence 15, Appl
40	86	6.0	6280	3 US-08-893-327-17	Sequence 17, Appl
41	86	6.0	6280	3 US-08-893-327-19	Sequence 19, Appl
42	86	6.0	6367	1 US-08-470-299-1	Sequence 1, Appl
43	86	6.0	6367	3 US-08-778-511-3	Sequence 3, Appl
44	86	6.0	6926	1 US-08-470-299-2	Sequence 2, Appl
45	86	6.0	7106	1 US-08-188-281B-8	Sequence 8, Appl

ALIGNMENTS

RESULT 1
US-08-345-913-1
Sequence 1, Application US/08345913
Patent No. 5641665
GENERAL INFORMATION:
APPLICANT: Hobart, Peter
APPLICANT: Parker, Suzanne
APPLICANT: Margalith, Michal
APPLICANT: Khatibi, Shirin
TITLE OF INVENTION: PLASMIDS SUITABLE FOR IL-2 EXPRESSION
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Knobbe, Martens, Olson and Bear
STREET: 620 Newport Center Drive 16th Floor
CITY: Newport Beach
STATE: CA
COUNTRY: USA
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: Fast-Seq Version 1.5
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/345,913
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Ways Vensko, Nancy
REGISTRATION NUMBER: 36,298
REFERENCE/DOCKET NUMBER: VICAL.043A
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619-235-8550
TELEFAX: 619-235-0176
TELEX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 4928 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE:
ORIGINAL SOURCE:
FEATURE:

RESULT 3
US-09-132-808-1
; Sequence 1, Application US/09132808
; Patent No. 6197332
; GENERAL INFORMATION:
; APPLICANT: Ronald Zuckermann et

```

1  RESULT      4
2  US-08-910-647-2
3  ; sequence 2, Application US/08910647
4  ; Patent No. 6251433
5  ;
6  GENERAL INFORMATION:
7  ;
8  ; APPLICANT: Zuckermann et al.
9  ;
10 ; TITLE OF INVENTION: Compositions and Methods for
11 ;                    Polynucleotide Delivery
12 ;
13 ; NUMBER OF SEQUENCES: 4
14 ;
15 ; CORRESPONDENCE ADDRESS:
16 ; ADDRESS: Chiron Corporation
17 ; STREET: 4560 Horton Street
18 ; CITY: Emeryville
19 ; STATE: California
20 ; COUNTRY: U.S.A.
21 ;
22 ; ZIP: 94608-2916
23 ;
24 ; COMPUTER READABLE FORM:
25 ; MEDIUM TYPE: Floppy disk
26 ;
27 ; COMPUTER: IBM PC compatible
28 ;
29 ; OPERATING SYSTEM: PC-DOS/MS-DOS
30 ;
31 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
32 ;
33 ; CURRENT APPLICATION DATA:
34 ; APPLICATION NUMBER: US/08/910,647
35 ; FILING DATE:

```

```

; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-910-647-2

Query Match          7.3%; Score 104.8; DB 4; Length 4328;
Best Local Similarity 94.0%; Pred. No. 1.2e-23;
Matches 109; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 tgagcagtaactcgttgctgcgcgcgcgcaccacagacataatagctgacagactaacaga 60
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1546 TGAGCAGTACTCGTTGCTGCGCGCGCGCCACCACAGACATAATAGCTGACAGACTAACAGA 1605

QY 61 ctgttcctttccatggtgcgcgtctttttctgcagtcacgcgtctgcagactgtgatcaga 116
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1606 CTGTTCCTTTCCATGGGTCTTTTCTGTCAGTCACCGCTGCTGCACCTTAAGAATTGAGA 1661

```

```

RESULT      5
US-08-910-647-4
; Sequence 4, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-4

```

```

Query Match          7.3%; Score 104.8; DB 4; Length 4818;
Best Local Similarity 94.0%; Pred. NO. 1.3e-23;
Matches 109; Conservative 0; Mismatches 7; Indels 0; Gaps
QY      1 tgagcagtactggtctgcgcgcgcacccaccagacataatagctacagactaacaga 60
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1549 TGAGCAGTACTGTTGCTGCCCGGGCCGCCACCACACATAATAGCTACAGACTTAACAGA 1608
QY      61 ctgttccttccatgggtcttttctgcagtcaccgctcgacacgctgtgatcaga 116
        ||| | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      1609 CTGTTCCTTTCCATGGTGCTTTTTCTGCAGTCACCGCTCGTCGACCTAAGAATTTCAGA 1664

RESULT      6
US-08-910-647-3
; Sequence 3, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910.647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-3

```

```

Query Match      7.3%; Score 104.8; DB 4; Length 5107;
Best Local Similarity 94.0%; Pred. NO. 1.4e-23;
Matches 109; Conservative 0; Mismatches 7; Indels 0; Gaps

QY      1  tgcagcagtactggttctgcgcgcgcgcacacagacataatagctgcacagactaacaga 60
Db      1549  TGAGCAGTACTCGTTGCTCGCGCGCGCCACCACACATAATAGCTGCACAGACTAACAGA 1608

QY      61  ctgtctcttccatagggtcttttgcagtcaccgtctgcacgcggtgatcaga 116
Db      1609  CTGTTCCCTTTCATGGGTCTTTTCTGTCAGTCACTGCTGTCGACCTAAGAATTTCAGA 1664

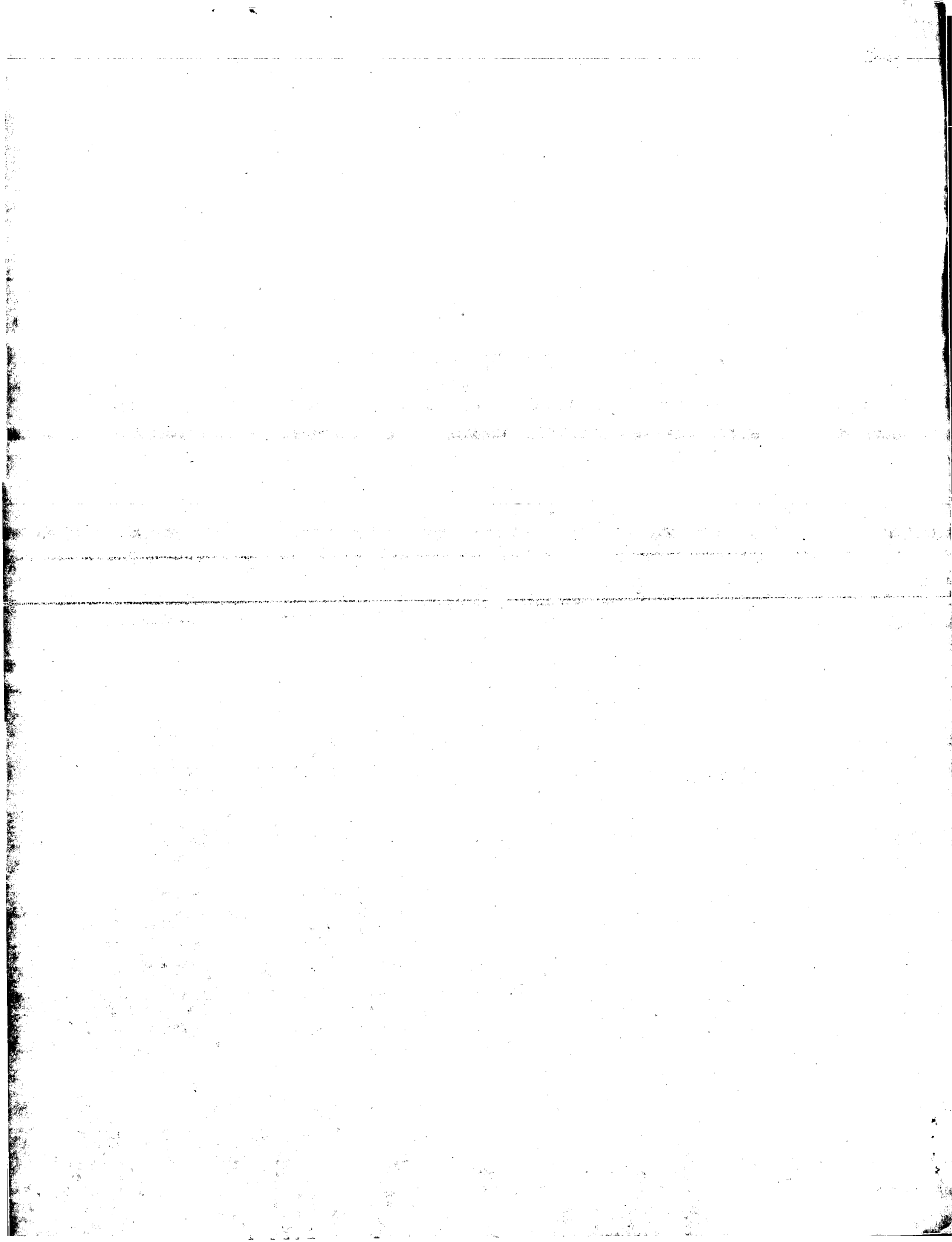
RESULT      7
US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION

```

```

;
; STRANDEDNESS: single
; TOPOLOGY: linear
;

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:29 ; Search time 7419.31 Seconds
(without alignments)
2065.351 Million cell updates/sec

Title: US-09-391-606-12
Perfect score: 1426
Sequence: 1 tggacgtactctgtgtgc.....tcctgacctggaaggtgc 1426

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST: *
1: em_estfun: *
2: em_esthum: *
3: em_estin: *
4: em_estom: *
5: em_estpl: *
6: em_estba: *
7: em_estro: *
8: em_estov: *
9: em_hic: *
10: gb_est1: *
11: gb_est2: *
12: gb_hic: *
13: gb_gss: *
14: em_gss_fun: *
15: em_gss_hum: *
16: em_gss_inv: *
17: em_gss_pln: *
18: em_gss_pro: *
19: em_gss_rod: *
20: em_gss_vrt: *
21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	70	4.9	149	10	AW669676
2	70	4.9	162	10	AW669522
3	70	4.9	162	10	AW785253
4	70	4.9	162	10	AW785901
5	70	4.9	243	10	AW689518
6	70	4.9	328	10	AW669634
7	70	4.9	374	10	AW669633
8	70	4.9	503	11	BF074689
9	54	3.8	98	11	BG365021
10	42.2	3.0	566	13	AQ778622
11	41.6	2.9	337	11	BF712368
12	41.6	2.9	359	11	BF711512

C 13	41.6	2.9	359	11	BF712352	MI-P-A3-a
C 14	41.6	2.9	396	11	BF712329	MI-P-A3-a
C 15	41.6	2.9	403	11	BF702314	MI-P-A2-a
C 16	41.6	2.9	432	11	BF711501	MI-P-A2-a
C 17	41.6	2.9	448	11	BF712292	MI-P-A3-a
C 18	41.6	2.9	448	11	BF712400	MI-P-A3-a
C 19	41.6	2.9	528	11	BF712343	MI-P-A3-a
C 20	41.6	2.9	530	11	BF703911	MI-P-A3-a
C 21	41.6	2.9	534	11	BF712419	MI-P-A3-a
C 22	41.6	2.9	537	11	BF702273	MI-P-A2-a
C 23	40.2	2.8	593	11	BF263367	HV_CEA000
C 24	40	2.8	143	10	AW359614	46376 MAR
C 25	40	2.8	294	11	BF199164	249242 MA
C 26	40	2.8	294	11	BF199165	249243 MA
C 27	40	2.8	358	11	BF711604	MI-P-A1-a
C 28	40	2.8	359	11	BF703303	MI-P-A3-a
C 29	40	2.8	363	11	BF702121	MI-P-A1-a
C 30	40	2.8	404	11	BF711339	MI-P-A1-a
C 31	40	2.8	432	11	BF712300	MI-P-A3-a
C 32	40	2.8	433	11	BG894486	355131 MA
C 33	40	2.8	439	11	BF712448	MI-P-A3-a
C 34	40	2.8	456	11	BF712424	MI-P-A3-a
C 35	40	2.8	527	11	BF701994	MI-P-A1-a
C 36	40	2.8	532	11	BF702078	MI-P-A1-a
C 37	40	2.8	535	11	BF702139	MI-P-A1-a
C 38	40	2.8	537	11	BF702127	MI-P-A1-a
C 39	40	2.8	575	11	BF712372	MI-P-A3-a
C 40	40	2.8	582	11	BF198279	248040 MA
C 41	35.8	2.8	1101	13	CNS0039G	AL063921
C 42	35.2	2.7	675	13	AZ840016	2M0136D20
C 43	39	2.7	690	10	AU034107	AU034107
C 44	38.4	2.7	1321	11	BG490965	602519759
C 45						

ALIGNMENTS

RESULT 1
AW669676 149 bp mRNA EST 25-APR-2001
LOCUS 113173 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION AW669676
ACCESSION AW669676
VERSION AW669676.1 GI:7526190
KEYWORDS EST.
SOURCE BOS taurus
COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 149)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T., Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G., Pertea,G., Holt,I., Karameycheva,S., Liang,F., Quackenbush,J. and Keefe,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
JOURNAL Genome Res. 11 (4), 626-630 (2001)
MEDLINE 21180013
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail@marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCAGCAG
Plate: 107 row: D column: 22

Seq primer: ATTAGGTGACACTATAG.

FEATURES

source

Location/Qualifiers

1. .149

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 1BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

24 a 48 c 38 g 39 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity 100.0%; Pred. No. 7.5e-10; Length 149;

Matches: 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 ctgtgcttctagtgcagccatctgtgttggccctcccccgtgcttcttgacc 1416

Db 36 CTGTGCTTCTAGTGCAGCCATCTGTGTTCGCCCTCCCGTCCCTTCCTTGACCC 95

QY 1417 tggaaaggtgc 1426

Db 96 TGAAGGTGC 105

RESULT 2

AW669522

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

MEDLINE

COMMENT

TITLE

JOURNAL

MEDLINE

COMMENT

FEATURES

source

Location/Qualifiers

1. .162

/organism="Bos taurus"

/db_xref="taxon:9913"

/clone_lib="MARC 1BOV"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from lymph node, ovary,

fat, hypothalamus, and pituitary."

24 a 48 c 38 g 39 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Query Match

Best Local Similarity

Matches

QY 1357

Db 49

QY 1417

tggaaaggtgc

1426

fat, hypothalamus, and pituitary."

28 a 51 c 42 g 41 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Query Match

Best Local Similarity

Matches

QY 1357

Db 49

QY 1417

tggaaaggtgc

1426

Db 109

TGAAGGTGC

118

RESULT 3

AW785253

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

FEATURES

source

Location/Qualifiers

1. .162

/organism="Sus scrofa"

/db_xref="taxon:9823"

/clone_lib="MARC 1PIG"

/tissue_type="pooled"

/lab_host="DH10B"

/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;

Library made from pooled tissue from day 11, 13, 15, 20,

and 30 embryos."

28 a 51 c 42 g 41 t

BASE COUNT

ORIGIN

Query Match

Best Local Similarity

Matches

Query Match

Best Local Similarity

Matches

QY 1357

Db 49

QY 1417

tggaaaggtgc

1426

```

|||||
Db 109 TGAAGGTGC 118

RESULT 4
AW785901
LOCUS 162 bp mRNA EST 09-JUL-2000
DEFINITION 117399 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW785901
VERSION AW785901.1 GI:7842677
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 162)
Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
EST discovered in use of two pooled tissue normalized cDNA libraries for
EST discovered in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGCG
Plate: 37 row: K column: 23
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..162
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 28 a 51 c 42 g 41 t
ORIGIN
|||||
Query Match 4.9%; Score 70; DB 10; Length 162;
Best Local Similarity 100.0%; Pred. No. 7.8e-10;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 cgtgccccttctagttgcagccatctgtttgttgcctcccccgccttcccttgcacc 1416
|||||
Db 49 cgtgccccttctagttgcagccatctgtttgttgcctcccccgccttcccttgcacc 108
|||||
QY 1417 tggaaaggtgc 1426
|||||
Db 109 TGAAGGTGC 118

RESULT 5
AW669518
LOCUS 243 bp mRNA EST 25-APR-2001
DEFINITION 119291 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW669518
VERSION AW669518.1 GI:7526032
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
1 (bases 1 to 328)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 243)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGCG
Plate: 106 row: J column: 22
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..243
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 56 a 67 c 59 g 61 t
ORIGIN
|||||
Query Match 4.9%; Score 70; DB 10; Length 243;
Best Local Similarity 100.0%; Pred. No. 9e-10;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1357 cgtgccccttctagttgcagccatctgtttgttgcctcccccgccttcccttgcacc 1416
|||||
Db 49 cgtgccccttctagttgcagccatctgtttgttgcctcccccgccttcccttgcacc 108
|||||
QY 1417 tggaaaggtgc 1426
|||||
Db 109 TGAAGGTGC 118

RESULT 6
AW669634
LOCUS 328 bp mRNA EST 25-APR-2001
DEFINITION 113105 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW669634
VERSION AW669634.1 GI:7526148
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
Bovidae; Bovinae; Bos.
1 (bases 1 to 328)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

```

Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 107 row: I column: 14
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..328
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC IBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
71 a 88 c 101 g 58 t
BASE COUNT
ORIGIN
Query Match 4.9%; Score 70; DB 10; Length 328;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1357 ctgtgccttctagtgcagccatctgttggccctcccccgtgcttccttgacc 1416
Db 110 CTGTGCTTCTAGTTCAGCAGCATCTGTTGTGCCCCCTCCCGCTTCTCTTGACCC 51
QY 1417 tgaaggtgc 1426
Db 50 TGAAGGTGC 41
RESULT 7
AW669633/c 374 bp mRNA EST 25-APR-2001
LOCUS 113104 MARC IBOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION AW669633
ACCESSION AW669633.1 GI:7526147
VERSION EST.
KEYWORDS
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 374)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 107 row: I column: 14
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..328
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC IBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
84 a 102 c 109 g 79 t
BASE COUNT
ORIGIN
Query Match 4.9%; Score 70; DB 10; Length 374;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1357 ctgtgccttctagtgcagccatctgttggccctcccccgtgcttccttgacc 1416
Db 109 CTGTGCTTCTAGTTCAGCAGCATCTGTTGTGCCCCCTCCCGCTTCTCTTGACCC 50
QY 1417 tgaaggtgc 1426
Db 49 TGAAGGTGC 40
RESULT 8
BF074689 503 bp mRNA EST 25-APR-2001
LOCUS 222166 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BF074689
ACCESSION BF074689.1 GI:10868200
VERSION EST.
KEYWORDS
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 503)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 82 row: E column: 4
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..503
/organism="Bos taurus"
/db_xref="taxon:9913"

and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 107 row: I column: 13
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..374
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC IBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
84 a 102 c 109 g 79 t
BASE COUNT
ORIGIN
Query Match 4.9%; Score 70; DB 10; Length 374;
Best Local Similarity 100.0%; Pred. No. 1e-09;
Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1357 ctgtgccttctagtgcagccatctgttggccctcccccgtgcttccttgacc 1416
Db 109 CTGTGCTTCTAGTTCAGCAGCATCTGTTGTGCCCCCTCCCGCTTCTCTTGACCC 50
QY 1417 tgaaggtgc 1426
Db 49 TGAAGGTGC 40
RESULT 8
BF074689 503 bp mRNA EST 25-APR-2001
LOCUS 222166 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
DEFINITION BF074689
ACCESSION BF074689.1 GI:10868200
VERSION EST.
KEYWORDS
SOURCE COW.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 503)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perte, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 82 row: E column: 4
Seq primer: ATTAGGTGACACTATAG.
Location/Qualifiers
1..503
/organism="Bos taurus"
/db_xref="taxon:9913"

/clone_lib="MARC 2BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from testis, thymus,
 semitendinosus muscle, longissimus muscle, pancreas,
 adrenal, and endometrium."
 BASE COUNT 122 a 146 c 130 g 105 t
 ORIGIN

Query Match 4.9%; Score 70; DB 11; Length 503;
 Best Local Similarity 100.0%; Pred. No. 1.2e-09;
 Matches 70; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1357 cgtgaccttgaagcagccatgtgtttgccccctcccccgccttcccttgaccc 1416
 |||||
 Db 367 CTTGCTCTAGTTCAGCCAGCCATCTGTTGTTGCCCCCTCCCGTCCCTCTGACCC 426
 QY 1417 tggaaggtgc 1426
 |||||
 Db 427 TGAAGGTGC 436

RESULT 9
 LOCUS BG365021 98 bp mRNA EST 25-APR-2001
 DEFINITION 100689 MARC 1BOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION BG365021
 VERSION BG365021.1 GI:13254118
 KEYWORDS EST.
 SOURCE cow.
 ORGANISM Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
 Bovidae; Bovinae; Bos.
 1 (bases 1 to 98)
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
 Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
 G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
 Perte, G., Holt, I., Karameycheva, S., Liang, F., Quackenbush, J. and
 Keele, J.W.
 Sequence evaluation of four pooled-tissue normalized bovine cDNA
 libraries and construction of a gene index for cattle
 Genome Res. 11 (4), 626-630 (2001)
 21180013
 Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smith@email.marc.usda.gov
 Single pass sequencing. Bases called and alt_trimmed with phred
 v0.980904 e. Vector identified by cross_match with the 'minscore 18
 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAACAGCTATGACCAT
 BACKWARD: GTTTCACGTCACGACG
 Plate: 72 row: L column: 16
 Seq primer: ATTTAGGTGACACTATAG.
 Location/Qualifiers
 1..98
 /organism="Bos taurus"
 /db_xref="taxon:9913"
 /clone_lib="MARC 1BOV"
 /tissue_type="pooled"
 /lab_host="DH10B"
 /note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
 Library made from pooled tissue from lymph node, ovary,
 fat, hypothalamus, and pituitary."
 BASE COUNT 18 a 35 c 18 g 27 t
 ORIGIN

FEATURES
 source
 1..566
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="Plate-2235 Col=16 Row=C"
 /note="Organ: sperm; Vector: pheloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 160 a 101 c 101 g 201 t 3 others
 ORIGIN

Query Match 3.0%; Score 42.2; DB 13; Length 566;
 Best Local Similarity 51.6%; Pred. No. 0.21;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 596 tgttgaaatttacacagacacctcttctcttggtggcgtagcgctgtggagccttatg 655
 |||||
 Db 248 TTTTAAATATTCAGAAATCAAAATTTTTCACAAAGTAATATCTATGAGAGTTAAG 189
 QY 656 ggaatgcggtgtgtgcaactttgggagctggaattccaatatgcacagtcaccaacctaagt 715
 |||||
 Db 188 AGTATGTGAAGTGAATTAAGCTGTATATGTACTACACAGCACCAAGATCAGACATAGTT 129
 QY 716 tgaagaacttaagtgtatctgtaacgtatcgcaattctctgttaaacacacccaaggccta 775
 |||||
 Db 128 CTAACAATAACTGCATATGTTGACTTAGTGCACACTAGTAGTAAATTAAGCCTATATANA 69
 QY 776 taata 779

Query Match 3.8%; Score 54; DB 11; Length 98;
 Best Local Similarity 100.0%; Pred. No. 3.6e-05;
 Matches 54; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1373 ccagccatctgtgttggccccctcccccgccttcccttgacccgtggaaggtgc 1426
 |||||
 Db 1 CCAGCCATCTGTTGTTGCCCCCTCCCGTGCCTTCTTGACCTGGAAGGTGC 54

RESULT 10
 LOCUS AQ778622/c 566 bp DNA GSS 02-AUG-1999
 DEFINITION HS.2235_A2_B08_T7C CIT Approved Human Genomic Sperm Library D Homo
 sapiens genomic clone Plate=2235 Col=16 Row=C, DNA sequence.
 ACCESSION AQ778622
 VERSION AQ778622.1 GI:5681582
 KEYWORDS GSS.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 566)
 Mahairas, G.G., Wallace, J.C., Smith, K., Swartzell, S., Holzman, T.,
 Keller, A., Shaker, R., Furlong, J., Young, J., Zhao, S., Adams, M.D. and
 Hood, L.
 Sequence-tagged connectors: A sequence approach to mapping and
 scanning the human genome
 Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
 99380589
 Contact: Mahairas GG, Wallace JC, Hood L
 High Throughput Sequencing Center
 University of Washington
 401 Queen Anne Avenue North, Seattle, WA 98109, USA
 Tel: (206) 616-3618
 Fax: (206) 616-3887
 Email: jwallace@u.washington.edu
 Clones may be purchased from Research Genetics (info@resgen.com).
 BAC end Web Server: http://www.htsc.washington.edu
 Plate: 2235 row: C column: 16
 Seq primer: T7
 Class: BAC ends
 High quality sequence stop: 566.
 Location/Qualifiers
 1..566
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pheloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 160 a 101 c 101 g 201 t 3 others
 ORIGIN

Query Match 3.0%; Score 42.2; DB 13; Length 566;
 Best Local Similarity 51.6%; Pred. No. 0.21;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 596 tgttgaaatttacacagacacctcttctcttggtggcgtagcgctgtggagccttatg 655
 |||||
 Db 248 TTTTAAATATTCAGAAATCAAAATTTTTCACAAAGTAATATCTATGAGAGTTAAG 189
 QY 656 ggaatgcggtgtgtgcaactttgggagctggaattccaatatgcacagtcaccaacctaagt 715
 |||||
 Db 188 AGTATGTGAAGTGAATTAAGCTGTATATGTACTACACAGCACCAAGATCAGACATAGTT 129
 QY 716 tgaagaacttaagtgtatctgtaacgtatcgcaattctctgttaaacacacccaaggccta 775
 |||||
 Db 128 CTAACAATAACTGCATATGTTGACTTAGTGCACACTAGTAGTAAATTAAGCCTATATANA 69
 QY 776 taata 779

FEATURES
 source
 1..566
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pheloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 160 a 101 c 101 g 201 t 3 others
 ORIGIN

Query Match 3.0%; Score 42.2; DB 13; Length 566;
 Best Local Similarity 51.6%; Pred. No. 0.21;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 596 tgttgaaatttacacagacacctcttctcttggtggcgtagcgctgtggagccttatg 655
 |||||
 Db 248 TTTTAAATATTCAGAAATCAAAATTTTTCACAAAGTAATATCTATGAGAGTTAAG 189
 QY 656 ggaatgcggtgtgtgcaactttgggagctggaattccaatatgcacagtcaccaacctaagt 715
 |||||
 Db 188 AGTATGTGAAGTGAATTAAGCTGTATATGTACTACACAGCACCAAGATCAGACATAGTT 129
 QY 716 tgaagaacttaagtgtatctgtaacgtatcgcaattctctgttaaacacacccaaggccta 775
 |||||
 Db 128 CTAACAATAACTGCATATGTTGACTTAGTGCACACTAGTAGTAAATTAAGCCTATATANA 69
 QY 776 taata 779

FEATURES
 source
 1..566
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="CIT Approved Human Genomic Sperm Library D"
 /sex="male"
 /note="Organ: sperm; Vector: pheloBAC11; BAC Clones in
 E-Coli DH10B"
 BASE COUNT 160 a 101 c 101 g 201 t 3 others
 ORIGIN

Query Match 3.0%; Score 42.2; DB 13; Length 566;
 Best Local Similarity 51.6%; Pred. No. 0.21;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 596 tgttgaaatttacacagacacctcttctcttggtggcgtagcgctgtggagccttatg 655
 |||||
 Db 248 TTTTAAATATTCAGAAATCAAAATTTTTCACAAAGTAATATCTATGAGAGTTAAG 189
 QY 656 ggaatgcggtgtgtgcaactttgggagctggaattccaatatgcacagtcaccaacctaagt 715
 |||||
 Db 188 AGTATGTGAAGTGAATTAAGCTGTATATGTACTACACAGCACCAAGATCAGACATAGTT 129
 QY 716 tgaagaacttaagtgtatctgtaacgtatcgcaattctctgttaaacacacccaaggccta 775
 |||||
 Db 128 CTAACAATAACTGCATATGTTGACTTAGTGCACACTAGTAGTAAATTAAGCCTATATANA 69
 QY 776 taata 779

Query Match 3.0%; Score 42.2; DB 13; Length 566;
 Best Local Similarity 51.6%; Pred. No. 0.21;
 Matches 95; Conservative 0; Mismatches 89; Indels 0; Gaps 0;
 QY 596 tgttgaaatttacacagacacctcttctcttggtggcgtagcgctgtggagccttatg 655
 |||||
 Db 248 TTTTAAATATTCAGAAATCAAAATTTTTCACAAAGTAATATCTATGAGAGTTAAG 189
 QY 656 ggaatgcggtgtgtgcaactttgggagctggaattccaatatgcacagtcaccaacctaagt 715
 |||||
 Db 188 AGTATGTGAAGTGAATTAAGCTGTATATGTACTACACAGCACCAAGATCAGACATAGTT 129
 QY 716 tgaagaacttaagtgtatctgtaacgtatcgcaattctctgttaaacacacccaaggccta 775
 |||||
 Db 128 CTAACAATAACTGCATATGTTGACTTAGTGCACACTAGTAGTAAATTAAGCCTATATANA 69
 QY 776 taata 779

Db 68 TAAA 65

RESULT 11
BF712368/c

LOCUS
DEFINITION MI-P-A3-adf-d-06-1-UM.s1 MI-P-A3 Sus scrofa cDNA clone
MI-P-A3-adf-d-06-1-UM 3', mRNA sequence.

ACCESSION
VERSION BF712368.1 GI:12011843

KEYWORDS
SOURCE EST.

ORGANISM
Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.

REFERENCE
AUTHORS 1 (bases 1 to 337)
TITLE Normalization and subtraction: two approaches to facilitate gene

JOURNAL
MEDLINE
COMMENT

Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized anterior pituitary at estrus day 12 library cDNA
Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal
Science Research Center, Department of Animal Science, University
of Missouri-Columbia, 65211 Clone distribution: clones will be
available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..337

/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-A3-adf-d-06-1-UM"
/clone_lib="MI-P-A3"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-A3
library is derived from anterior pituitary at estrus day
12. For a detailed description of the library from which
this clone was derived, please visit our web site at
http://pigest.genome.iastate.edu/.

TAG_LIB=MI-P-A3
TAG_TISSUE=anterior pituitary at estrus day 12
TAG_SEQ=TGACAG

BASE COUNT 70 a 81 c 106 g 80 t

Query Match 2.9%; Score 41.6; DB 11; Length 337;
Best Local Similarity 84.3%; Pred. No. 0.27;
Matches 59; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1357 ctgtgccttcctagttgcagccatctgtgttgcctcccccgccttccttgacc 1416
|||||

Db 133 CTGTGCTTCTAGTTGCTGGCATCTCT--GTTGCCCTCCCGCAGTACCTCCCTTGACCC 76
|||||

QY 1417 tgaaggtgc 1426
|||||

Db 75 TGAAGAGTGC 66

RESULT 12
BF711512/c

LOCUS
DEFINITION MI-P-A2-acq-e-09-1-UM.s1 MI-P-A2 Sus scrofa cDNA clone
MI-P-A2-acq-e-09-1-UM 3', mRNA sequence.

ACCESSION
VERSION BF711512.1 GI:12010987

KEYWORDS
SOURCE EST.

ORGANISM
Sus scrofa

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 359)
Bonaldo, M.F., Lennon, G. and Soares, M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery

Genome Res. 6 (9), 791-806 (1996)
97044477

Contact: Tuggle CK
Molecular Genetics Laboratory, Department of Animal Science
Iowa State University
201 Kildee Hall, Ames, IA 50011-3150, USA

Tel: 5152944252
Fax: 5152942401
Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the
oligonucleotide that was used to prime the synthesis of first
strand cDNA and therefore this may represent a bonafide poly A
tail. The sequence tag present in the cDNA between the NotI site
and the oligo-dT track served to verify it as a clone from the
non-normalized anterior pituitary at estrus day 5 library cDNA
Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal
Science Research Center, Department of Animal Science, University
of Missouri-Columbia, 65211 Clone distribution: clones will be
available through Research Genetics (www.resgen.com)
Seq primer: M13 Forward
POLYA=Yes.

FEATURES
source

Location/Qualifiers
1..359

/organism="Sus scrofa"
/strain="crossbreed"
/db_xref="taxon:9823"
/clone="MI-P-A2-acq-e-09-1-UM"
/clone_lib="MI-P-A2"
/lab_host="DH10B (Life Technologies)"
/note="Vector: pT7T3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: EcoRI; The MI-P-A2
library is derived from anterior pituitary at estrus day
5. For a detailed description of the library from which
this clone was derived, please visit our web site at
http://pigest.genome.iastate.edu/.

TAG_LIB=MI-P-A2
TAG_TISSUE=anterior pituitary at estrus day 5
TAG_SEQ=TCGCGT

BASE COUNT 68 a 91 c 108 g 92 t

Query Match 2.9%; Score 41.6; DB 11; Length 359;
Best Local Similarity 84.3%; Pred. No. 0.27;
Matches 59; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1357 ctgtgccttcctagttgcagccatctgtgttgcctcccccgccttccttgacc 1416
|||||

Db 133 CTGTGCTTCTAGTTGCTGGCATCTCT--GTTGCCCTCCCGCAGTACCTCCCTTGACCC 76
|||||

QY 1417 tgaaggtgc 1426
|||||

Db 75 TGAAGAGTGC 66

RESULT 13

BF712352/c
 LOCUS
 DEFINITION MI-P-A3-adf-b-07-1-UM.sl MI-P-A3 Sus scrofa cDNA clone
 ACCESSION BF712352
 VERSION MI-P-A3-adf-b-07-1-UM 3', mRNA sequence.
 KEYWORDS EST.
 SOURCE BF712352.1 GI:12011827
 ORGANISM pig.
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 359)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT 9704477
 Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized anterior pituitary at estrus day 12 library cDNA
 Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal
 Science Research Center, Department of Animal Science, University
 of Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
source

Location/Qualifiers

1..359
 /organism="Sus scrofa"
 /strain="crossbreed"
 /db_xref="taxon:9823"
 /clone_lib="MI-P-A3-adf-b-07-1-UM"
 /lab_host="MI-P-A3"
 /note="Vector: pT7T3D-pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-A3
 library is derived from anterior pituitary at estrus day
 12. For a detailed description of the library from which
 this clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/
 TAG_LIB=MI-P-A3
 TAG_TISSUE=anterior pituitary at estrus day 12
 TAG_SEQ=TGACAG"

BASE COUNT 71 a 91 c 106 g 91 t
 ORIGIN
 Query Match 2.9%; Score 41.6; DB 11; Length 359;
 Best Local Similarity 84.3%; Pred. No. 0.27;
 Matches 59; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1357 cgtgctcttagtgccagccatctgtttgtgccccccccgccttccttgaccc 1416
 |||||
 Db 133 CTGTGCTTCTAGTCTGGGATCTCT--GTTGCCCTCCAGTACCTCCTTGACCC 76

QY 1417 tgaaggtgc 1426
 |||||
 Db 75 TGAAGGTGC 66

RESULT 14

BF712329/c
 LOCUS
 DEFINITION MI-P-A3-adf-b-04-1-UM.sl MI-P-A3 Sus scrofa cDNA clone
 ACCESSION BF712329
 VERSION MI-P-A3-adf-b-04-1-UM 3', mRNA sequence.
 KEYWORDS EST.
 SOURCE BF712329.1 GI:12011804
 ORGANISM pig.
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 396)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT 9704477
 Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: cktuggle@iastate.edu

The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. The sequence tag present in the cDNA between the NotI site
 and the oligo-dT track served to verify it as a clone from the
 non-normalized anterior pituitary at estrus day 12 library cDNA
 Library Preparation: RJ Woods, JA Green, RS Prather S142 Animal
 Science Research Center, Department of Animal Science, University
 of Missouri-Columbia, 65211 Clone distribution: clones will be
 available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=yes.

FEATURES
source

Location/Qualifiers

1..396
 /organism="Sus scrofa"
 /strain="crossbreed"
 /db_xref="taxon:9823"
 /clone_lib="MI-P-A3-adf-b-04-1-UM"
 /lab_host="MI-P-A3"
 /note="Vector: pT7T3D-pac (Pharmacia) with a modified
 polylinker; Site 1: Not I; Site 2: EcoRI; The MI-P-A3
 library is derived from anterior pituitary at estrus day
 12. For a detailed description of the library from which
 this clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/
 TAG_LIB=MI-P-A3
 TAG_TISSUE=anterior pituitary at estrus day 12
 TAG_SEQ=TGACAG"

BASE COUNT 77 a 103 c 118 g 98 t
 ORIGIN
 Query Match 2.9%; Score 41.6; DB 11; Length 396;
 Best Local Similarity 84.3%; Pred. No. 0.28;
 Matches 59; Conservative 0; Mismatches 9; Indels 2; Gaps 1;

QY 1357 cgtgctcttagtgccagccatctgtttgtgccccccccgccttccttgaccc 1416
 |||||
 Db 133 CTGTGCTTCTAGTCTGGGATCTCT--GTTGCCCTCCAGTACCTCCTTGACCC 76

QY 1417 tgaaggtgc 1426
 |||||
 Db 75 TGAAGGTGC 66

RESULT 15

BF702314/c
 LOCUS
 DEFINITION MI-P-A2-aar-f-03-1-UM.sl MI-P-A2 Sus scrofa cDNA clone
 ACCESSION 3F702314
 VERSION MI-P-A2-aar-f-03-1-UM 3', mRNA
 KEYWORDS EST
 SOURCE 3F702314
 ORGANISM pig.
 Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 403)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 JOURNAL discovery
 MEDLINE Genome Res. 6 (9), 791-806 (1996)
 COMMENT 9704477
 Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: cktuggle@iastate.edu

MI-P-A2-aar-f-03-1-UM 3', mRNA sequence.

ACCESSION BF702314
 VERSION BF702314.1 GI:11987722
 KEYWORDS EST.
 SOURCE pig.
 ORGANISM Sus scrofa
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
 REFERENCE 1 (bases 1 to 403)
 AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
 TITLE Normalization and subtraction: two approaches to facilitate gene
 discovery
 JOURNAL Genome Res. 6 (9), 791-806 (1996)
 MEDLINE 97044477
 COMMENT Contact: Tuggle CK
 Molecular Genetics Laboratory, Department of Animal Science
 Iowa State University
 201 Kildee Hall, Ames, IA 50011-3150, USA
 Tel: 5152944252
 Fax: 5152942401
 Email: cktuggle@iastate.edu
 The sequence contained an oligo-dT track that was present in the
 oligonucleotide that was used to prime the synthesis of first
 strand cDNA and therefore this may represent a bonafide poly A
 tail. cDNA Library Preparation: RJ Woods, JA Green, RS Prather S142
 Animal Science Research Center, Department of Animal Science,
 University of Missouri-Columbia, 65211 Clone distribution: clones
 will be available through Research Genetics (www.resgen.com)
 Seq primer: M13 Forward
 POLYA=Yes.

FEATURES

source location/Qualifiers
 1..403
 /organism="Sus scrofa"
 /strain="crossbreed"
 /db_xref="taxon:9823"
 /clone="MI-P-A2-aar-f-03-1-UM"
 /clone_lib="MI-P-A2"
 /lab_host="DH10B (Life Technologies)"
 /note="Vector: pT73b-Pac (Pharmacia) with a modified
 polylinker; Site_1: Not 1; Site_2: EcoRI; The MI-P-A2
 library is derived from anterior pituitary at estrus day
 5. For a detailed description of the library from which
 this clone was derived, please visit our web site at
 http://pigest.genome.iastate.edu/
 TAG_SEQ=None found"
 BASE COUNT 79 a 105 c 119 g 100 t
 ORIGIN

Query Match 2.9%; Score 41.6; DB 11; Length 403;
 Best Local Similarity 84.3%; Pred. NO. 0.28; Mismatches 9; Indels 2; Gaps 1;
 Matches 59; Conservative 0;
 QY 1357 ctgtgccttctagtgcagccatctgtgttgcctcccccgtgccttccttgaccc 1416
 |||||
 Db 131 CTGTGCTTCTAGTCTGGCATCTCT--GTTGCCCTCCCCAGTACCTCCCTTGACCC 74
 |||||
 QY 1417 tggaaagtgc 1426
 |||||
 Db 73 TGGAAAGTGC 64

Search completed: February 7, 2002, 21:32:32
 Job time: 20666 sec

```

Query Match      6.5%   Score 77.5;   DB 18;   Length 3672;
Best Local Similarity 22.2%;   Pred. No. 1.3e+02;
Matches 41;   Conservative 15;   Mismatches 50;   Indels 79;   Gaps 7;

QY  31 NTEVPSTSTTEKAATNWKYKVCVWOLWKGKHSQVP-----WINGOKKPLYL 78
      |||||
db 1585 niflpsk-----wqlvitylqqpqgkrrthgkislwsggrkpdvt 1626

```

'D 198

1707

DE H. pylori outer membrane polypeptide.
 XX Outer membrane polypeptide; OMP; vaccine; H. pylori infection; humoral;
 KW cellular immune response.
 XX Helicobacter pylori.
 OS WO9921959-A2.
 XX 06-MAY-1999.
 XX 28-OCT-1998; 98WO-US22883.
 XX 17-DEC-1997; 97US-0993001.
 XX 28-OCT-1997; 97US-0959131.
 XX (GENO-) GENOME THERAPEUTICS CORP.
 XX Alm RA, Ellis RW, Guild BC, Noonan BM, Smith D;
 PI WPI; 1999-326698/27.
 XX N-PSDB; AAX75836.
 XX Cellular vaccine against Helicobacter pylori
 PT Claim 7; Page 311-312; 352pp; English.
 XX The invention relates to a vaccine for preventing or treating infections
 CC by Helicobacter pylori. The vaccine contains at least one isolated
 CC H. pylori polypeptide, or its fragments, in a carrier, where the
 CC carrier is a Salmonella, Vibrio cholerae or Shigella vector containing a
 CC nucleic acid encoding the H. pylori polypeptide. The vaccines induce
 CC humoral and cellular immune responses. The vaccines are used to treat or
 CC prevent infections by H. pylori. Sequences AAX75779 to AAX75837 represent
 CC nucleic acid sequences encoding H. pylori outer membrane polypeptides
 CC (OMPs) AAX17160 to AAX17218.
 XX Sequence 327 AA;
 SQ

Query Match 6.8%; Score 79; DB 20; Length 327;
 Best Local Similarity 25.7%; Pred. No. 3;
 Matches 58; Conservative 23; Mismatches 71; Indels 74; Gaps 17;

QY 48 AMKYKVC-VQWLVGKHGSOVPHN-CQKKPLYLYGAFLLMPLAKATKTTLNGKENLAWFI 105
 DB 118 amtnyvgwglvg---qkpwscglatgnltqggy--dkmtqgmtqisnqeala--- 169
 QY 106 GGTGLG----GLRKAGDWSATVRYEVEALSV-----PEIDVSGIGRG-----NL--L 146
 DB 170 astcaktayadhrnyviynayirnykdifeirgryespadymsynggldmtlnlgnf 229
 QY 147 KFW----FAQAJAAN-----YDPKE---ANSFT-----NYKGSALYMYGIT 181
 DB 230 kfwfssfgfaynewlynfyspkytlnqgtinpgvhafyilwnykgfs-----iq 283
 QY 182 DLSLFRAYGAY-----SPANDKLGSDFTFR-KFDLGIISAF 217
 DB 284 pfvyrspfnedpnftdydsnptftgig----frsqtdvtvlnpf 325

RESULT 12
 AAY96159
 ID AAY96159 standard; Protein; 753 AA.
 XX AAY96159;
 XX 19-DEC-2000 (first entry)
 XX Arabidopsis thaliana OPT family member emb CAB43855.1.
 XX Oligopeptide transporter; OPT; isp4-like protein; OPT1; enkephalin;
 KW oploid.

XX Arabidopsis thaliana.
 OS WO200052162-A2.
 XX 08-SEP-2000.
 XX 01-MAR-2000; 2000WO-US05158.
 XX 01-MAR-1999; 99US-0122312.
 XX (UYTE-) UNIV TENNESSEE RES CORP.
 XX (BECK/) BECKER J M.
 XX (HAUS/) HAUSER M.
 XX (DONH/) DONHARDT A.
 XX (BARN/) BARNES D.
 XX Becker JM, Hauser M, Donhardt A, Barnes D;
 PI WPI; 2000-587311/55.
 XX Obtaining mammalian enkephalin transport proteins for transporting
 PT enkephalin across cell membrane, comprises transforming OPT1 gene
 PT deleted yeast strain with mammalian gene library and selecting the
 PT cells -
 XX Disclosure; Fig 5; 59pp; English.
 XX The present sequence is that of an oligopeptide transporter (OPT)
 CC family member (isp4-like protein) from Arabidopsis thaliana,
 CC designated emb CAB43855.1. Members of the OPT family have been
 CC identified in yeasts and plants. The invention relates to a novel
 CC OPT of Saccharomyces cerevisiae, designated OPT1 (see AAY96149).
 CC OPT1 is the first example of a genetically defined eukaryotic
 CC transport protein which can transport enkephalins across the cell
 CC membrane. Vectors allowing expression of the OPT1 gene, transgenic
 CC plants, and treatment of OPT1p with toxic enkephalins as an
 CC antifungal method are claimed. Also claimed is a method for
 CC obtaining mammalian enkephalin transporters by functional
 CC complementation of OPT1 deficient yeast.
 XX Sequence 753 AA;
 SQ

Query Match 6.7%; Score 78; DB 21; Length 753;
 Best Local Similarity 22.2%; Pred. No. 12;
 Matches 45; Conservative 16; Mismatches 50; Indels 92; Gaps 10;

QY 8 WV---VEGILN--RLPKQ-----FFVKCSVDVNTFVPSSETSTTEKAATNAMYKY 53
 DB 564 wllssvenicntdmlpksspwtpcpdvvfynasii-wgilpggrmftskgi-----ypg 616
 QY 54 CVWQWLVGKHGSOVP-----WINGOKKPLYLYGAFLLMPLAKATKTTLNGKENL 101
 DB 617 mmwffligflapvpwvffarkfpekkwhqihiplifsganvm-pmakavhy----- 667
 QY 102 AWFIGTGLGLRKAGDWSATVRYEVEALSVPEIDVSGIGRGNLLKFWFAOAIANYDPK 161
 DB 668 -----ws-----wfvagvifny--- 679

QY 162 EANSFTNYKGSALYMYGITDLSL 184
 DB 680 --yifrykgwwarhnyllsaal 700

RESULT 13
 AAW23596
 ID AAW23596 standard; Protein; 2001 AA.
 XX AAW23596;
 XX 10-FEB-1998 (first entry)
 XX

QY 79 YGAFMLNPLAKATKTTLNKENLAW-FIGGTLGG-----LRKAGDWSATVRVEYVEALSVP 133
Db 1627 ldfml-----prktslssdsnkftcmighclssqeeflqlagkwdl-----1667
QY 134 EIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGSALYMYGITDSLSFRAYGAYS 193
Db 1668 -----gnllllfngakvgsq-----eaflyacgpnhtsvmpckyg 1702
QY 194 KPAND 198
Db 1703 kfvnd 1707

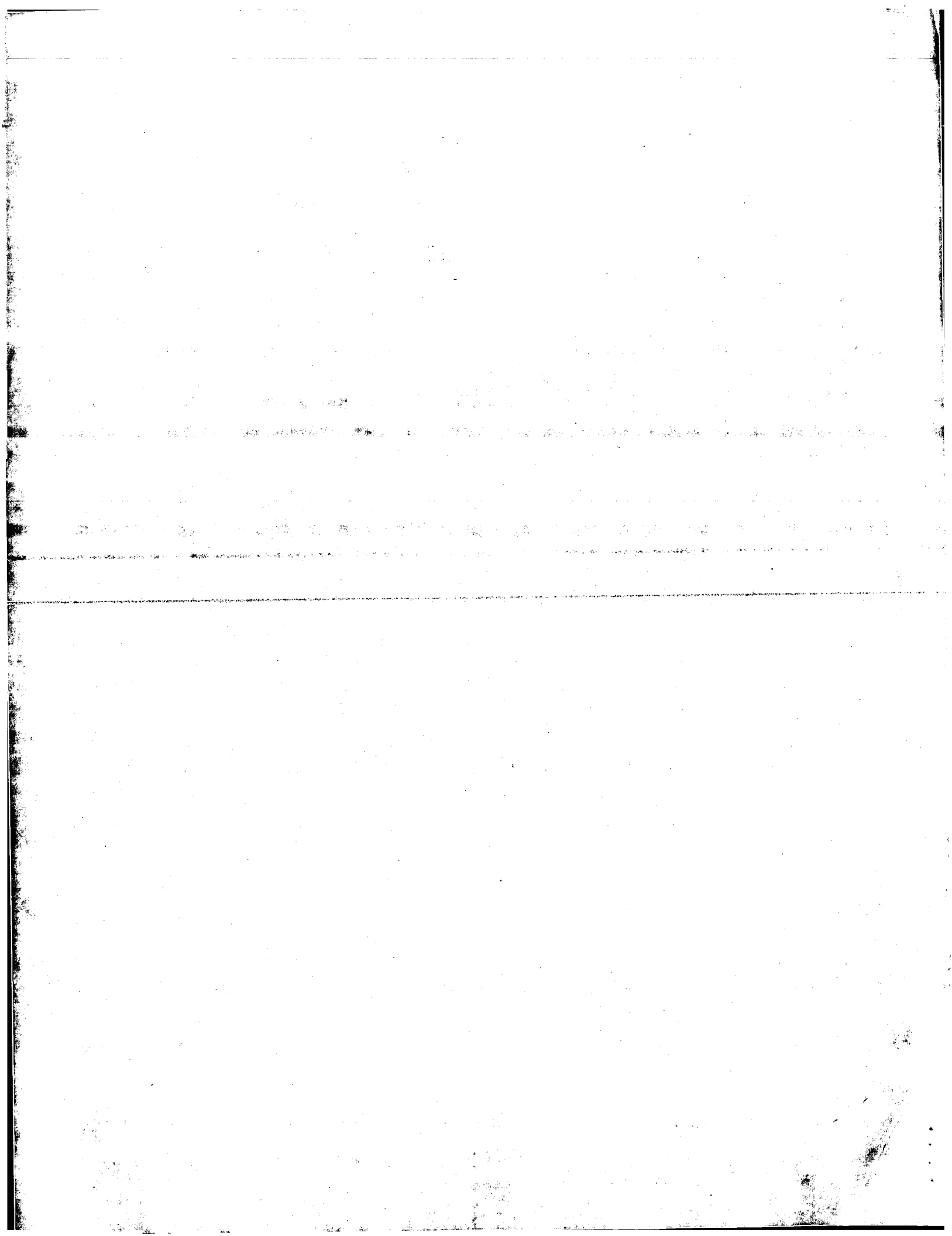
RESULT 15
AAW31949
ID AAW31949 standard; Protein; 3801 AA.
XX AAW31949;
AC AAW31949;
XX 14-APR-1998 (first entry)
XX Human bg protein associated with Chediak-Higashi syndrome.
XX Chediak-Higashi syndrome; beige; bg gene; intracellular vesicle;
KW human; therapy; diagnosis.
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 3694..4708
FT /note= "WD40 or G protein-beta subunit repeat motif"
FT Misc-difference 3767
FT /note= "encoded by TNC"
XX
XX WO9734914-A1.
PN
PD 25-SEP-1997.
XX
XX 21-MAR-1997; 97WO-US05068.
XX 01-JUL-1996; 96US-0021064.
XX 22-MAR-1996; 96US-0013883.
XX 19-APR-1996; 96US-0015673.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA (UTAH) UNIV UTAH.
XX
XX Kaplan J, Moore KJ, Perou CM;
PI
XX
XX WPI; 1997-502728/46.
DR N-PSDB; AAT89256.
XX
XX Murine and human long and short form beige genes - useful to screen
PT for compounds to treat, or to diagnose intracellular vesicle
PT disorders, specifically Chediak-Higashi syndrome
XX
XX Claim 9; Fig 7; 133pp; English.

This polypeptide is encoded by the novel human beige (bg) gene
(see AAT89256) that is involved in the normal differentiation and/or
function of intracellular vesicles. The amino acid sequence of
the predicted bg gene product indicates that the protein is novel.
An alternatively spliced transcript of the human bg gene (see
AAT89257) encodes a 3672-amino acid protein (see AAW31950). Mutation
of the human bg gene is responsible for human Chediak-Higashi
syndrome (CHS). An intracellular vesicle disorder, specifically
CHS, can be diagnosed by measuring bg gene expression in a patient
sample, e.g. by detecting mRNA transcripts of the bg gene, the bg
gene product or a bg gene mutation contained in the genome of the
mammal, specifically in a splice site of the bg gene (claimed).
Genetically engineered host cells can be used to screen for

CC compounds useful for the treatment of intracellular vesicle
CC disorders, specifically CHS, e.g. a molecule that modulates the
CC expression of the bg gene in a mammal (claimed).
XX Sequence 3801 AA;
SQ

Query Match 6.6%; Score 77.5; DB 18; Length 3801;
Best Local Similarity 22.2%; Pred. No. 1.4e+02;
Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;
QY 31 NTFVPSSETTEKAATNAMYKYCYVMQWLVGKHSQVP-----WINGQKKPLYL 78
Db 1585 niflpsk-----wqlvlvlylqpgqkrrrhgkislwvsgqrkpdvt 1626
QY 79 YGAFMLNPLAKATKTTLNKENLAW-FIGGTLGG-----LRKAGDWSATVRVEYVEALSVP 133
Db 1627 ldfml-----prktslssdsnkftcmighclssqeeflqlagkwdl-----1667
QY 134 EIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGSALYMYGITDSLSFRAYGAYS 193
Db 1668 -----gnllllfngakvgsq-----eaflyacgpnhtsvmpckyg 1702
QY 194 KPAND 198
Db 1703 kfvnd 1707

Search completed: February 7, 2002, 21:34:34
Job time: 20753 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:54:39 ; Search time 243.49 Seconds
(without alignments)
1024.076 Million cell updates/sec

Title: US-09-391-606-14
Perfect score: 1101
Sequence: 1 atgttgcttaggaacc.....tatctggtcagtcagattc 1101

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	37	3.4	40	US-09-025-176-2	Sequence 2, Appli
2	35.2	3.2	48	US-08-116-389-6	Sequence 6, Appli
3	35.2	3.2	48	US-08-708-431-6	Sequence 6, Appli
4	35.2	3.2	48	US-08-880-830-6	Sequence 6, Appli
5	35.2	3.2	48	PCT-US94-13895-6	Sequence 6, Appli
6	34.2	3.1	1978	US-08-137-175A-4	Sequence 4, Appli
7	34.2	3.1	1978	US-08-479-017-4	Sequence 4, Appli
8	34.2	3.1	3167	US-08-472-534-1	Sequence 1, Appli
9	34.2	3.1	4320	US-08-472-534-4	Sequence 1, Appli
10	33.2	3.0	490	US-08-245-511-1	Sequence 1, Appli
11	33.2	3.0	490	US-08-600-993A-1	Sequence 1, Appli
12	33.2	3.0	2019	US-08-245-511-46	Sequence 46, Appl
13	33.2	3.0	2019	US-08-600-993A-46	Sequence 46, Appl
14	33.2	2.9	1569	US-08-680-726A-57	Sequence 57, Appl
15	32.2	2.9	1569	US-09-092-409-57	Sequence 57, Appl
16	32.2	2.9	3263	US-08-714-918-20	Sequence 20, Appl
17	32.2	2.9	3263	US-09-265-315-20	Sequence 20, Appl
18	32.2	2.9	3263	US-09-265-315-20	Sequence 20, Appl
19	32.2	2.9	3263	US-09-266-417-20	Sequence 20, Appl
20	32.2	2.9	10592	US-08-680-726A-51	Sequence 51, Appl
21	32.2	2.9	10592	US-08-680-726A-51	Sequence 51, Appl
22	32.2	2.9	10592	US-09-092-409-51	Sequence 51, Appl
23	32.2	2.9	10592	US-09-092-409-51	Sequence 51, Appl
24	32	2.9	1017	US-08-714-918-45	Sequence 45, Appl
25	32	2.9	1017	US-09-265-315-45	Sequence 45, Appl
26	32	2.9	1017	US-09-265-315-45	Sequence 45, Appl
27	32	2.9	1017	US-09-266-417-45	Sequence 45, Appl

28	31.8	2.9	1005	4	US-09-446-504-79	Sequence 79, Appl
29	31.6	2.9	1307	2	US-08-867-030B-6	Sequence 6, Appli
30	31.6	2.9	1307	5	PCT-US95-06119-6	Sequence 6, Appli
31	31.2	2.8	2359	1	US-08-426-169-5	Sequence 5, Appli
32	31.2	2.8	2359	4	US-09-233-813-5	Sequence 5, Appli
33	31.2	2.8	2359	5	PCT-US95-09470-5	Sequence 5, Appli
C 34	31.2	2.8	3758	3	US-08-323-477-1	Sequence 1, Appli
C 35	31.2	2.8	9432	1	US-08-277-231A-1	Sequence 1, Appli
C 36	31.2	2.8	9432	2	US-08-473-750-4	Sequence 4, Appli
C 37	31.2	2.8	9432	2	US-08-477-326-4	Sequence 4, Appli
C 38	31	2.8	1717	1	US-08-229-515A-12	Sequence 12, Appl
C 39	31	2.8	1717	1	US-08-645-863-12	Sequence 12, Appl
C 40	31	2.8	3660	1	US-08-158-232-42	Sequence 42, Appl
C 41	31	2.8	3660	1	US-08-611-928-42	Sequence 42, Appl
C 42	31	2.8	3660	3	US-09-173-891-42	Sequence 42, Appl
C 43	30.8	2.8	618	4	US-08-953-326-22	Sequence 22, Appl
C 44	30.4	2.8	1425	3	US-09-009-494-5	Sequence 5, Appli
C 45	30.4	2.8	1425	4	US-09-010-232-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-09-025-176-2/c
; Sequence 2, Application US/09025176
; Patent No. 6258532
; GENERAL INFORMATION:
; APPLICANT: Charles W. Stratton
; APPLICANT: William M. Mitchell
; TITLE OF INVENTION: Methods for in vitro and in vivo
; TITLE OF INVENTION: susceptibility testing of Chlamydia species
; FILE REFERENCE: 50150/006003
; CURRENT APPLICATION NUMBER: US/09/025,176
; CURRENT FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: 08/911,593
; PRIOR FILING DATE: 1997-08-14
; NUMBER OF SEQ ID NOS: 2
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 40
; TYPE: DNA
; ORGANISM: Chlamydia pneumoniae
US-09-025-176-2

Query Match 3.4%; Score 37; DB 4; Length 40;
Best Local Similarity 100.0%; Pred. No. 0.0065;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1065 cgagagagctgtcagtcagtcagtcagtcagtcagtcagtc 1101
|||||
DB 40 CGAGAGAGCTGCTACGTCATCTGTCAGTTCAGATTC 4

RESULT 2
US-08-116-389-6
; Sequence 6, Application US/08116389
; Patent No. 5601978
; GENERAL INFORMATION:
; APPLICANT: Burczak, John
; APPLICANT: Carrino, J.J.
; APPLICANT: Salituro, J.A.
; APPLICANT: Pabich, E.K.
; APPLICANT: Klonowski, P.A.
; APPLICANT: Manlove, M.T.
; APPLICANT: Marshall, R.L.
; TITLE OF INVENTION: Materials and Methods for the Detection
; TITLE OF INVENTION: of Chlamydia Trachomatis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: One Abbott Park Road

us-09-391-606-14.rni

Fri Feb 8 08:44:54 2002

```

; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PCDOS/MSDOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/116,389
; FILING DATE: 03 SEPTEMBER 1993
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Brainard, Thomas D.
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5372.US.01
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708 937-4884
; TELEFAX: 708 938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 48 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Chlamydia trachomatis
; US-08-116-389-6

Query Match 3.2%; Score 35.2; DB 1; Length 48;
Best Local Similarity 83.3%; Pred. No. 0.028;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 318 ttggatcgcttggatgtttctgtacttttagagagcttctaaggtta 365
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 TTGGATCGTGTGGATGTATCTGTACATTAGGAGCCACCAGTGGATA 48

RESULT 4
US-08-880-830-6
; Sequence 6, Application US/08880830
; Patent No. 5846785
; GENERAL INFORMATION:
; APPLICANT: Burczak, John
; APPLICANT: Carrino, J.J.
; APPLICANT: Salituro, J.A.
; APPLICANT: Pabich, E.K.
; APPLICANT: Klonowski, P.A.
; APPLICANT: Manlove, M.T.
; APPLICANT: Marshall, R.L.
; TITLE OF INVENTION: Materials and Methods for the Detection
; of Chlamydia Trachomatis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PCDOS/MSDOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/880,830
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/449,863
; FILING DATE:
; APPLICATION NUMBER: 08/116,389
; FILING DATE: 03 SEPTEMBER 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Brainard, Thomas D.
; REGISTRATION NUMBER: 32,459
; REFERENCE/DOCKET NUMBER: 5372.US.01
; TELECOMMUNICATION INFORMATION:
; US-08-708-431-6

Query Match 3.2%; Score 35.2; DB 1; Length 48;
Best Local Similarity 83.3%; Pred. No. 0.028;
Matches 40; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

Qy 318 ttggatcgcttggatgtttctgtacttttagagagcttctaaggtta 365
||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 1 TTGGATCGTGTGGATGTATCTGTACATTAGGAGCCACCAGTGGATA 48

RESULT 3
US-08-708-431-6
; Sequence 6, Application US/08708431
; Patent No. 5756298
; GENERAL INFORMATION:
; APPLICANT: Burczak, John
; APPLICANT: Carrino, J.J.
; APPLICANT: Salituro, J.A.
; APPLICANT: Pabich, E.K.
; APPLICANT: Klonowski, P.A.
; APPLICANT: Manlove, M.T.
; APPLICANT: Marshall, R.L.
; TITLE OF INVENTION: Materials and Methods for the Detection
; of Chlamydia Trachomatis
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ABBOTT LABORATORIES
; STREET: One Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PCDOS/MSDOS
; SOFTWARE: WordPerfect
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/708,431
; FILING DATE: 05-SEP-1996
; CLASSIFICATION: 435
```


Fri Feb 8 08:44:54 2002

NAME/KEY: CDS
 LOCATION: 127..948
 OTHER INFORMATION: /product= "OspA"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 962..1861
 OTHER INFORMATION: /product= "OspB"
 US-08-137-175A-4

Query Match 3.1%; Score 34.2; DB 1; Length 1978;
 Best Local Similarity 48.2%; Pred. No. 0.47;
 Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 835 attgtcagccaaactacacagctgttttaaaacttaactgcagccttcatgcaaccctcttta 894
 DB 1715 AGTGTGACAGTAAACAACTAAAGATTTTGTGTTCTTAACAGATGGTACCATACAGTA 1774

QY 895 ctaggaaatgccacagcattgtctactactgattgttctcagacattcatgcaaatgtt 954
 DB 1775 CAAGCATATGACACAGCAGGTACTAACTTGAGGGCAACTCAAGTGAATTAAGATCTT 1834

QY 955 tctgtcagatcaacaagttaataatctagaaaagcttggagttactgtaggagctact 1014
 DB 1835 GCAGCACTTAAAGCTGCTTTAAATAACATAAAAGTAAACATCTTACATCGGCTAATACC 1894

QY 1015 ttagtgtgctgataaat 1033
 DB 1895 TTTGTAGGTGTTGTTATT 1913

RESULT 7
 US-08-479-017-4
 ; Sequence 4, Application US/08479017
 ; Patent No. 6143872
 ; GENERAL INFORMATION:
 ; APPLICANT: BARBOUR, Alan G.
 ; APPLICANT: BERGSTROEM, Sven
 ; APPLICANT: HANSSON, Lennart
 ; TITLE OF INVENTION: IMPROVEMENT IN BORRELIA BURGDORFERI AND
 ; TITLE OF INVENTION: PROPHYLAXIS
 ; NUMBER OF SEQUENCES: 22
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/479,017
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/137,175
 ; FILING DATE: 26-OCT-1993
 ; APPLICATION NUMBER: PC/US92/08972
 ; FILING DATE: 22-OCT-1992
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: COOPER, Iver P. 28, 005
 ; REGISTRATION NUMBER:
 ; REFERENCE/DOCKET NUMBER: BARBOUR-1B
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; TELEX: 248633
 ; INFORMATION FOR SEQ ID NO: 4:
 ; SEQUENCE CHARACTERISTICS:

LENGTH: 1978 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: cDNA
 ORIGINAL SOURCE:
 ORGANISM: Borrelia burgdorferi
 STRAIN: ACAL
 INDIVIDUAL ISOLATE: Swedish isolate, pt. acrodermatitis
 INDIVIDUAL ISOLATE: chronicum migrans
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 85..104
 OTHER INFORMATION: /function= "Primer"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 111..130
 OTHER INFORMATION: /function= "Primer"
 FEATURE:
 NAME/KEY: misc.feature
 LOCATION: 948..965
 OTHER INFORMATION: /function= "Primer"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 127..948
 OTHER INFORMATION: /product= "OspA"
 FEATURE:
 NAME/KEY: CDS
 LOCATION: 962..1861
 OTHER INFORMATION: /product= "OspB"
 US-08-479-017-4

Query Match 3.1%; Score 34.2; DB 3; Length 1978;
 Best Local Similarity 48.2%; Pred. No. 0.47;
 Matches 96; Conservative 0; Mismatches 103; Indels 0; Gaps 0;

QY 835 attgtcagccaaactacacagctgttttaaaacttaactgcagccttcatgcaaccctcttta 894
 DB 1715 AGTGTGACAGTAAACAACTAAAGATTTTGTGTTCTTAACAGATGGTACCATACAGTA 1774

QY 895 ctaggaaatgccacagcattgtctactactgattgttctcagacattcatgcaaatgtt 954
 DB 1775 CAAGCATATGACACAGCAGGTACTAACTTGAGGGCAACTCAAGTGAATTAAGATCTT 1834

QY 955 tctgtcagatcaacaagttaataatctagaaaagcttggagttactgtaggagctact 1014
 DB 1835 GCAGCACTTAAAGCTGCTTTAAATAACATAAAAGTAAACATCTTACATCGGCTAATACC 1894

QY 1015 ttagtgtgctgataaat 1033
 DB 1895 TTTGTAGGTGTTGTTATT 1913

RESULT 8
 US-08-472-534-1
 ; Sequence 1, Application US/08472534
 ; Patent No. 5919620
 ; GENERAL INFORMATION:
 ; APPLICANT: Hamel, Josee
 ; APPLICANT: Brodeur, Bernard R
 ; APPLICANT: Martin, Denis
 ; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
 ; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Fish & Neave
 ; STREET: 1251 Avenue of the Americas
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10020
 ; COMPUTER READABLE FORM:

```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,534
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: Biovac-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3167 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: 30..755
FEATURE:
NAME/KEY: CDS
LOCATION: 771..2912
OTHER INFORMATION: /product= *FuctI/HSP72 (C-16)
US-08-472-534-1

```

Query Match	3.1%;	Score	34.2;	DB	2;	Length	3167;
Best Local Similarity	49.2%;	Pred. No.	0.61;				
Matches	90;	Conservative	0;	Mismatches	93;	Indels	0;
							Gaps
QY	259	aataagcatttacacgatgcagagtggttcacataatcagcgtctcattgccttaaacatt	318				
Db	2914	AAGATGAGTCTATTGGATGAAGAGTATCTAAAAAATACAGAAAGTTTATAATGATT	2973				
QY	319	tgggatcgcttgatgtttctgtactttaggagcttctaattgggtaccattagaggaaac	378				
Db	2974	TGTAATCAAGCTGATAACTATAGAACAATCAAAAGATTTTATTGTAATAATATCCAATAGAA	3033				
QY	379	tctacagcgttcaactctcgttggtttatttcggagtttaaaaggtactactgtataatgcaaat	438				
Db	3034	TATTTAGCTAGATATAGAGAAATTTATATTAGCTGAGCATGATAGTTGTGTCAAAATGAT	3093				
QY	439	gaa	441				
Db	3094	GAA	3096				

RESULT 9
US-08-472-534-4
; Sequence 4, Application US/08472534
; Patent No. 5919620
; GENERAL INFORMATION:
; APPLICANT: Hamel, Jossee
; APPLICANT: Brodeur, Bernard R
; APPLICANT: Martin, Denis
; TITLE OF INVENTION: HEAT SHOCK PROTEIN HSP72 FROM
; TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Neave
; STREET: 1251 Avenue of the Americas
; CITY: New York

```

STATE: New York
COUNTRY: United States of America
ZIP: 10020
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/472,534
FILING DATE:
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Haley Jr, James F
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: Biovac-2
TELECOMMUNICATION INFORMATION:
TELEPHONE: 212-596-9000
TELEFAX: 212-596-9090
TELEX: 14-8367
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 4320 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
FEATURE:
NAME/KEY: CDS
LOCATION: 3265..4320
OTHER INFORMATION: /product= "NH2-terminal portion of
OTHER INFORMATION: DNA J"
FEATURE:
NAME/KEY: CDS
LOCATION: 682..2502
OTHER INFORMATION: /product= "Heat-Shock Protein 72"
FEATURE:
NAME/KEY: mat_peptide
LOCATION: 682..2502
US-08-472-534-4

Query Match 3.1%; Score 34.2; DB 2; Length 4320;
Best Local Similarity 49.2%; Pred. No. 0.73; Indels 0; Gaps 0;
Matches 90; Conservative 0; Mismatches 93;

QY 259 aataagcattacacgatgcagagtgttcactaatgcaggctctcatgtcttaaacatt 318
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2504 AAGATGAGTGATTGGATGAAGAGTATCTAAAAAATACAGCAAAAGTTTATAATGATTTT 2563
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 319 tgggtagctgttaagtgtttcttacttttaggagcttctaattaggttacattagaggaaac 378
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2564 TGTATCAAGCTGATACTATAGAACTCAAAAGATTTTATTGATATATTCCAATAGAA 2623
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 379 tctacagcgttcaatctcggttggtttatttcggaggttaaaaggttactactgtataatgcaaat 438
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2624 TATTGACTAGATATAGAGAAATATATTAGCTGAGCATGATAGTTCTGTGTCAAAATGAT 2683
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 439 gaa 441
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2684 GAA 2686

RESULT 10
US-08-245-511-1
; Sequence 1, Application US/08245511
; Patent No. 5928900
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert

```

TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THERON
NUMBER OF SEQUENCES: 59
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/600,993A
FILING DATE: 1-MAR-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 490 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
STRAIN: R6
IMMEDIATE SOURCE:
CLONE: SPRU98
FEATURE:
NAME/KEY: CDS
LOCATION: 1..490
US-08-600-993A-1

Query Match	3.08;	Score 33.2;	DB 2;	Length 490;
Best Local Similarity	61.6%;	Pred. NO. 0.45;	33;	Indels 0;
Matches 53;	Conservative 0;	Mismatches		Gaps 0;
Qy	795	acaatggctcgcagcaacttttgatgctgatacaatccgcatctgtccagccaaactacc	854	
Db	330	ACATCCCTGGAGCAACTTTAGGAGCTGATATGTCATTATTGTATTTCAACAACACTACA	389	
Qy	855	tacagctgttttaaaacttaactgcac	880	
Db	390	AAAAGACGAAGTAAACAATATTACAT	415	
RESULT 12				
US-08-245-511-46				
; Sequence 46, Application US/08245511				
; Patent No. 5928900				
; GENERAL INFORMATION:				
; APPLICANT: Masure, H Robert				

APPLICANT: Pearce, Barbara J
APPLICANT: Tuomanen, Elaine
TITLE OF INVENTION: BACTERIAL EXPORTED PROTEINS AND
TITLE OF INVENTION: ACCELLULAR VACCINES BASED THEREON
NUMBER OF SEQUENCES: 58
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/245,511
FILING DATE: 18-MAY-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/116,541
FILING DATE: 01-SEP-1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-069 CIP
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 490 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
STRAIN: R6
IMMEDIATE SOURCE:
CLONE: SPUR98
FEATURE:
NAME/KEY: CDS
LOCATION: 1..490
PS-08-245-511-1

```

Query Match      3.0%  Score 33.2;  DB 2;  Length 490;
Best Local Similarity  61.6%  Pred. No. 0.45;
Matches 53;  Conservative 0;  Mismatches 33;  Indels 0;  Gaps 0;

Qy  795  acaatggtctcgagaaattttgatgctgataacatccgcgcatgctccagccaaactacc 854
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  330  ACAATCCTTGGAGGACCACTTTAGAGCTGATATGTCATTATTGATATTCAACAACACTACA 389
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

Qy  855  tacagctgttttaaacttaactgcatt 880
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
Db  390  AAAAGACGAAGTAAACAATATTACAT 415
      |||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||

RESULT 11
US-08-600-993A-1
; Sequence 1, Application US/08600993A
; Patent No. 5981229
; GENERAL INFORMATION:
; APPLICANT: Masure, H Robert
; APPLICANT: Pearce, Barbara J
; APPLICANT: Tuomanen, Elaine

```


APPLICANT: Haanes, Elizabeth J.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/092,409
FILING DATE: 12-JUL-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 1569 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1566
US-09-092-409-57

Query Match 2.9%; Score 32.2; DB 3; Length 1569;
Best Local Similarity 43.7%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 142; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 140 atgttttcgacggtatctttaaagtagatgcacacatttctctatggagcgaagc 199
Db 107 ATGTTTATAGTACTACCCACCATTAAATGAGACTAAAAATTTACTAGAGGATGGGTAT 166
Qy 200 ctactggatccgctgctgcaaaactatactactcgcgtagatagaccccgccctaca 259
Db 167 TTTTACAAACCCGCCCTTCTTATTTGTAAGATGGGATATCAATATCTATATATGCTATG 226
Qy 260 ataagcatttacacgagtgagtggttcaactaagcaggttcattgccttaaacattt 319
Db 227 AACGTAATATTTGTAAGAGATATTTTGAATAAACGATGTAATATCTATATATGCTATG 286
Qy 320 gggatcgcttgatgtttctgtacttttagggagctcttaagttacattagaggaact 379
Db 287 ATTATCCCTAGCTAGCAGATTTTGAGATTAGTATTAATACTATTAATAAATAATATG 346
Qy 380 ctacagcgttcaactcgtttgtttatcgaggttaaggtactactgtaaatgcaaatg 439
Db 347 ATGTTTATTTTGTAAATGATAGTGTTCCTCAATAATACTACATAAATAAAGTGTATCC 406
Qy 440 aactaccacaaacgtttcttaagtaa 464
Db 407 ATATCACAAATGCTACTATAATAA 431

APPLICANT: Frank, Rexann S.
APPLICANT: Frank, Rexann S.
TITLE OF INVENTION: RECOMBINANT CANINE HERPESVIRUSES
NUMBER OF SEQUENCES: 92
CORRESPONDENCE ADDRESS:
ADDRESSEE: Sheridan Ross & McIntosh
STREET: 1700 Lincoln Street, Suite 3500
CITY: Denver
STATE: Colorado
COUNTRY: U.S.A.
ZIP: 80203
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/680,726A
FILING DATE: 12-JUL-1996
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Connell, Gary J.
REGISTRATION NUMBER: 32,020
REFERENCE/DOCKET NUMBER: 2618-46-C1
TELEPHONE: (303) 863-9700
TELEFAX: (303) 863-0223
INFORMATION FOR SEQ ID NO: 57:
SEQUENCE CHARACTERISTICS:
LENGTH: 1569 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1566
US-08-680-726A-57

Query Match 2.9%; Score 32.2; DB 1; Length 1569;
Best Local Similarity 43.7%; Pred. No. 1.8; Indels 0; Gaps 0;
Matches 142; Conservative 0; Mismatches 183; Indels 0; Gaps 0;

Qy 140 atgttttcgacggtatctttaaagtagatgcacacatttctctatggagcgaagc 199
Db 107 ATGTTTATAGTACTACCCACCATTAAATGAGACTAAAAATTTACTAGAGGATGGGTAT 166
Qy 200 ctactggatccgctgctgcaaaactatactactcgcgtagatagaccccgccctaca 259
Db 167 TTTTACAAACCCGCCCTTCTTATTTGTAAGATGGGATATCAATATCTATATATGCTATG 226
Qy 260 ataagcatttacacgagtgagtggttcaactaagcaggttcattgccttaaacattt 319
Db 227 AACGTAATATTTGTAAGAGATATTTTGAATAAACGATGTAATATCTATATATGCTATG 286
Qy 320 gggatcgcttgatgtttctgtacttttagggagcttctaagttacattagaggaact 379
Db 287 ATTATCCCTAGCTAGCAGATTTTGAGATTAGTATTAATACTATTAATAAATAAATG 346
Qy 380 ctacagcgttcaactcgtttgtttatcgaggttaaggtactactgtaaatgcaaatg 439
Db 347 ATGTTTATTTTGTAAATGATAGTGTTCCTCAATAATACTACATAAATAAAGTGTATCC 406
Qy 440 aactaccacaaacgtttcttaagtaa 464
Db 407 ATATCACAAATGCTACTATAATAA 431

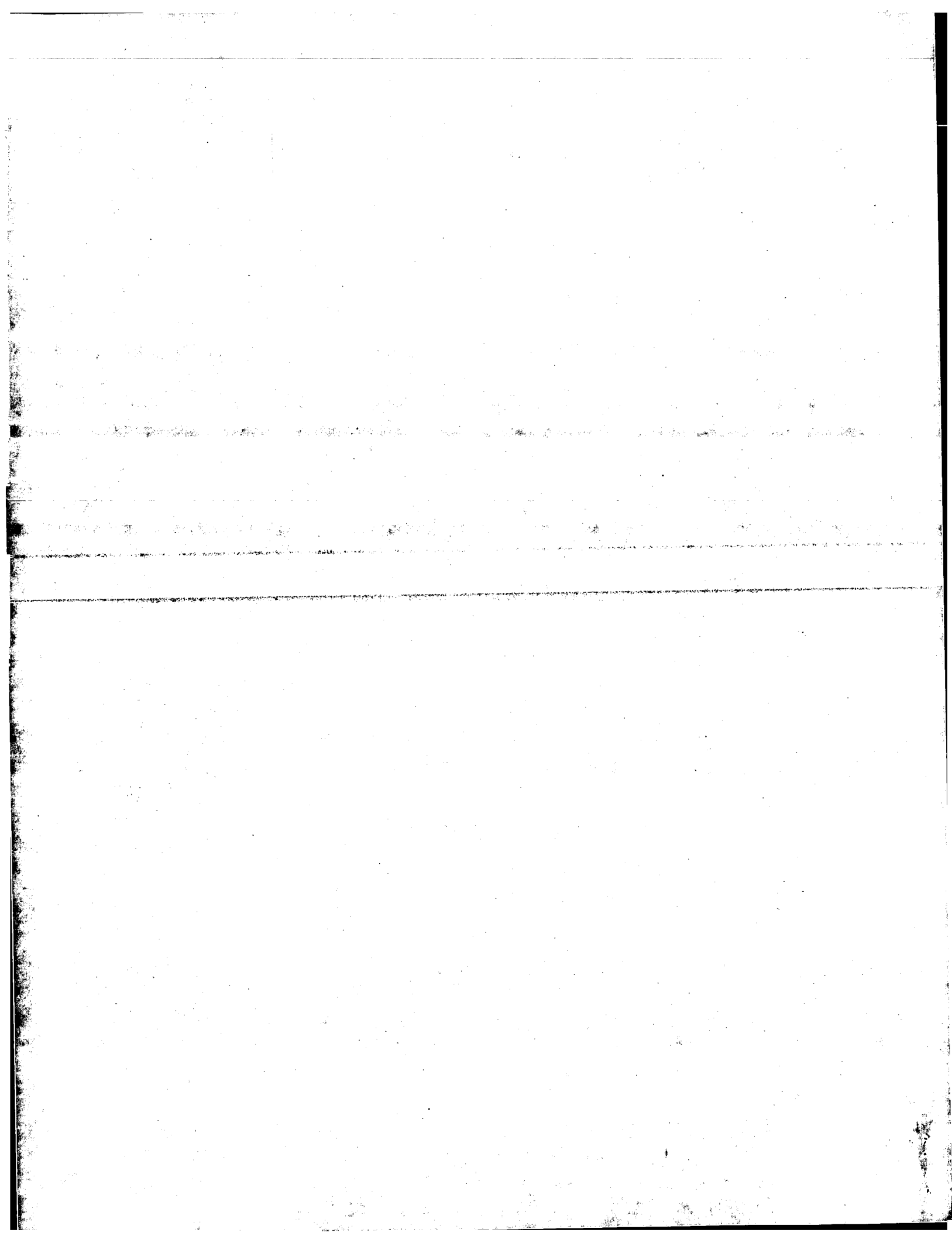
RESULT 15
US-09-092-409-57
; Sequence 57, Application US/09092409
; Patent No. 6159478
; GENERAL INFORMATION:

Fri Feb 8 08:44:54 2002

us-09-391-606-14.rni

Page 9

Search completed: February 7, 2002, 15:54:46
Job time: 400 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:32 ; Search time 7419.31 Seconds
(without alignments)
1594.636 Million cell updates/sec

Title: US-09-391-606-14
Perfect score: 1101
Sequence: 1 atgttgctgtaggaaccc.....tatctggtcagttcagattc 1101

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST:
1: em_estfun:
2: em_esthum:
3: em_estin:
4: em_estom:
5: em_estpl:
6: em_estba:
7: em_estro:
8: em_estov:
9: em_htc:
10: gb_est1:
11: gb_est2:
12: gb_htc:
13: gb_gss:
14: em_gss_fun:
15: em_gss_hum:
16: em_gss_inv:
17: em_gss_pln:
18: em_gss_pro:
19: em_gss_rtd:
20: em_gss_vrt:
21: em_gss_other:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	42.2	3.8	566	13	AQ778622 HS_2235_A
2	39.8	3.6	1101	13	CNS0039G
3	39.2	3.6	675	13	AZ840016
4	39	3.5	690	10	AU034107
5	38.4	3.5	1321	11	BG490965
6	38.2	3.5	296	10	AA381424
7	38.2	3.5	517	10	AW675282
8	38.2	3.5	534	10	AA431622
9	38	3.5	593	11	BF263367
10	37.6	3.4	869	13	BH131620
11	37	3.4	682	13	AQ925118
12	37	3.4	877	13	AZ671757

13	35.4	3.3	584	10	BE544037
14	36.2	3.3	539	13	AQ818532
15	36.2	3.3	1015	13	AZ686763
16	36.2	3.3	1041	13	CNS078LL
17	36	3.3	1101	13	CNS00LOO
18	35.8	3.3	418	11	BG153882
19	35.8	3.3	489	11	BI323992
20	35.8	3.3	749	13	AQ935992
21	35.8	3.3	852	13	CNS06229
22	35.8	3.3	887	13	AZ538399
23	35.4	3.2	274	10	BB424700
24	35.4	3.2	338	11	D71816
25	35.4	3.2	402	13	AQ477797
26	35.4	3.2	529	10	AW291247
27	35.2	3.2	254	10	BB081082
28	35.2	3.2	897	13	AZ677008
29	35.2	3.2	993	10	AL528462
30	35.2	3.2	994	13	AZ683097
31	35.2	3.2	1039	13	CNS07CAV
32	35.2	3.2	1101	13	CNS00FDD
33	35	3.2	310	10	AV277883
34	35	3.2	504	13	AZ617558
35	35	3.2	544	10	AW745780
36	35	3.2	771	11	BG584308
37	34.8	3.2	386	13	AQ442864
38	34.8	3.2	617	11	BE918310
39	34.8	3.2	700	10	AL503431
40	34.8	3.2	700	10	AL505428
41	34.8	3.2	786	11	BG218690
42	34.8	3.2	803	11	BG417399
43	34.8	3.2	985	10	BE412873
44	34.8	3.2	1041	13	CNS07BMO
45	34.6	3.1	371	13	AQ134473

ALIGNMENTS

RESULT 1
AQ778622/C
LOCUS: 566 bp DNA GSS 02-AUG-1999
DEFINITION: HS_2235_A2_B08_T7C CIT Approved Human Genomic Sperm Library D Homo sapiens genomic clone Plate=2235 Col=16 Row=C, DNA sequence.
ACCESSION: AQ778622
VERSION: AQ778622.1 GI:5681582
KEYWORDS: GSS.
SOURCE: human.
ORGANISM: Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE: 1 (bases 1 to 566)
AUTHORS: Mahairas,G.G., Wallace,J.C., Smith,K., Swartzell,S., Holzman,T., Keller,A., Shaker,R., Furlong,J., Young,J., Zhao,S., Adams,M.D., and Hood,L.
TITLE: Sequence-tagged connectors: A sequence approach to mapping and scanning the human genome
JOURNAL: Proc. Natl. Acad. Sci. U. S. A. 96 (17), 9739-9744 (1999)
MEDLINE: 99380589
COMMENT: Contact: Mahairas GG, Wallace JC, Hood L
High throughput Sequencing Center
University of Washington
401 Queen Anne Avenue North, Seattle, WA 98109, USA
Tel: (206) 616-3618
Fax: (206) 616-3887
Email: jwallace@u.washington.edu
Clones may be purchased from Research Genetics (info@resgen.com).
BAC end Web Server: http://www.htsc.washington.edu
Plate: 2235 row: C column: 16
Seq primer: T7
Class: BAC ends
High quality sequence stop: 566.
Location/Qualifiers
1..566

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 517)
 NIH-MGC <http://mgc.nci.nih.gov/>.
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished (1999)

using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) *Entamoeba histolytica*: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."

	Query Match	3.4%;	Score 37.6;	DB 13;	Length 869;
	Best Local Similarity	51.8%;	Prod. No. 7;		
	Matches 85;	Conservative 0;	Mismatches 79;	Indels 0;	Gaps 0;
Qy	861	tggtttaaacttaactgcataaggaacccctctttactcagaaatgccacagcattgtctac	920		
Dbb	45	TGTTTTTCAATAGGTCATATAAACCTATTATTATTATTATTCATTATTCATT	104		
Qy	921	tactgattcgtcttcagacttcactgcacaaattgtttctctgcagatccaaagtttaaatc	980		
Dbb	105	TAGATGTATGAAACAAGAATTAATGAAAGAATAAAAAATAAAACAACATGTTTATTTC	164		
Qy	981	tagaaaaagcttgytgattactctagagagctacttgattgatg	1024		
Dbb	165	CATAAAAATATATTTCGGGTATTAAACAATAATGACATGATG	208		

, DNA sequence.
 A0925118
 VERSION A0925118.1 GI:6614121
 KEYWORDS GSS.
 SOURCE house mouse.
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 1 (bases 1 to 682)
 REFERENCE Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret
 AUTHORS B., Levins, N., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P.
 and Fraser, C.M.
 TITLE Mouse BAC End Sequences from Library RPCI-23
 JOURNAL Unpublished (1999)

```

FEATURES
  source
    1. 682
      Location/Qualifiers
        Seq primer: T7
        Class: BAC ends.
        /organism="Mus musculus"
        /strain="C57BL/6J"
        /db_xref="taxon:10090"
        /clone="RPCI-23-302M22"
        /clone_lib="RPCI-23"

```


/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3 6; Site_1:
ECORI; Site_2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
ECORI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 270 a 121 c 117 g 172 t 2 others
ORIGIN

Query Match 3.4%; Score 37; DB 13; Length 682;
Best Local Similarity 49.7%; Pred. No. 9.6;
Matches 91; Conservative 0; Mismatches 92; Indels 0; Gaps 0;
QY 559 gaattccaatatgcacagtcacaaactaaagtgaagaacttaattgtatctgtcaactga 618
Db 447 GTATACCTTACATGATCCCAAAATGTACCGGGAACTCCTACAGCTGATAAAGAGCTTC 506
QY 619 tcgcaattctctgtaaacaccccaagggtataaaggcgtttgtttcccttgccaaca 678
Db 507 ACAAGTGGCTGAATACAAATTAACCTTAATAAATCACTAGCCTTCTCTATACAACA 566
QY 679 gacgctggctagcacagctactggaacaaagtctggaccatcaattatcatgaatgg 738
Db 567 CTCTNTACAAATGACCATGAATATGTAATAAATGTACTGGAGTAACTNTAACCAAGAGT 626
QY 739 caa 741
Db 627 GAA 629

RESULT 12
LOCUS A2671757 877 bp DNA GSS 14-DEC-2000
DEFINITION ENTKE65TR Entamoeba histolytica Sheared DNA Entamoeba histolytica
genomic, DNA sequence.
ACCESSION A2671757
VERSION A2671757.1 GI:11808903
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 877)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: entae@igr.org
Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
DNA library
Seq primer: M13-Reverse
Class: shotgun
High quality sequence start: 12
High quality sequence stop: 703.
Location/Qualifiers
1..877
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: pBACE3.6; Site_1: Bst I; Constructed at The
Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica
using a method described by Clark and Diamond (Clark,

C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
method for isolate identification. Exp. Parasitol.
77:450.). The DNA was mechanically sheared to give a
tight size distribution (~2 kb). The v + i method used for
the library construction is described in detail in Smith,
H.O. and Venter, J.C. (Making small insert libraries for
whole genome shotgun sequencing projects. In Genome
Sequencing: A Practical Approach, eds. M. Vaudin and B.
Bartell, Oxford University Press, 1999)."
BASE COUNT 232 a 99 c 182 g 364 t
ORIGIN

Query Match 3.4%; Score 37; DB 13; Length 877;
Best Local Similarity 53.0%; Pred. No. 10;
Matches 79; Conservative 0; Mismatches 70; Indels 0; Gaps 0;
QY 295 gcaggttcattgccttaaacatttggatgcgtttgtttctgtacttttaggagct 354
Db 585 GTATCTTATTTTTCATCAACTTTTGGAGTTTGTGTTTCAATTTTAAAGAGTT 644
QY 355 tctaatggttacattagaggaaactctacagcgttcactcgttggtttattcggagctt 414
Db 645 GTTAAGGTTCAACTTTTAGAGAAATATATCAGTGTAGATGTTCTTCTTCTTTTGA 704
QY 415 aaagtactactgtaaatgcacaaatgaact 443
Db 705 GTTGTTCATGTAATAATTAATAAAT 733

RESULT 13
LOCUS BE544037 584 bp mRNA EST 09-AUG-2000
DEFINITION 60106987E1 NIH_MGC_12 Homo sapiens cDNA clone IMAGE:3456213 5',
mRNA sequence.
ACCESSION BE544037
VERSION BE544037.1 GI:9772682
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 584)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: ATCC
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: Incyte Genomics, Inc.
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LILAM8443 row: n column: 22
High quality sequence stop: 577.
Location/Qualifiers
1..584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3456213"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 194 a 87 c 95 g 200 t 8 others
ORIGIN

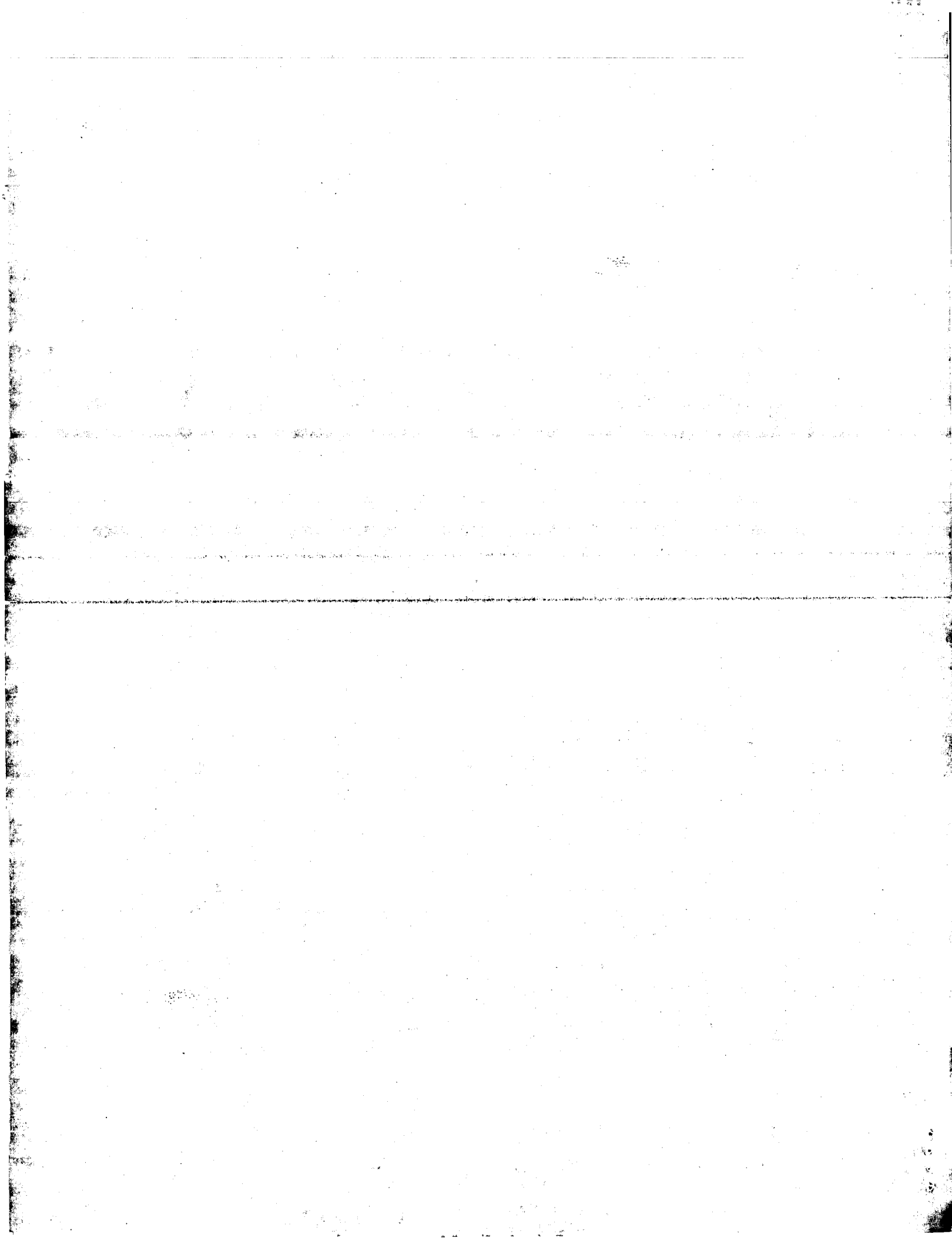
FEATURES
source
1..584
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:3456213"
/clone_lib="NIH_MGC_12"
/tissue_type="cervical carcinoma cell line"
/lab_host="DH10B"
/note="Organ: cervix; Vector: pCMV-SPORT6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.4 kb. Library prepared by Life
Technologies."
BASE COUNT 194 a 87 c 95 g 200 t 8 others
ORIGIN

BASE COUNT	pBACE3.6 vector at ECORI sites"			
	139 a	115 c	131 g	144 t
ORIGIN				
		Query Match	3.3%;	Score 36.2; DB 13; Length 539;
		Best Local Similarity	49.2%;	Pred. No. 15;
		Matches	95; Conservative	0; Mismatches 98; Indels 0; Gaps 0;
QY	857	cagctgttttaaccttaactgcatggaccctcttcttactagaaatgccacagcatgt	916	
Db	116	CAGTTGGCGTGACTCTATTTGGCAGGAGCATAAAGCCCTGGTCAATTTCTTACGGGTCAA	175	
QY	917	ctactactgattggtctcagacttcatgcataatgtttctctgcagatcaacaagtta	976	
Db	176	AAGAGACTGATTTGGTTTCATGCCCTCATGTCATGATGATGAGCCCTCACATCTGGTCTT	235	
QY	977	aattctagaaagcttgtggagtctactgaggagctacttagttgatgctgataaaatggt	1036	
Db	236	AATCTAAGATTCTGTGCTTCTCTGTAAGTCCCAATGGGTTTCATCTTGCCACATGCC	295	
QY	1037	cacttactgcaga	1049	
Db	296	CAGATAGAGCTGA	308	

```

RESULT 15
A2686763
LOCUS      A2686763      1015 bp      DNA      GSS      14-DEC-2000
DEFINITION      ENT1568TF Entamoeba histolytica Sheared DNA Entamoeba histolytica
                genomic, DNA sequence.
ACCESSION      A2686763.1      GI:11823909
VERSION        A2686763
KEYWORDS
SOURCE
ORGANISM      Entamoeba histolytica.
                Eukaryota; Entamoebidae; Entamoeba.
REFERENCE
AUTHORS      Loftus, B., Van Aken, S. and Fraser, C.
TITLE        Determination of clone end sequences from Entamoeba histolytica
                HMI:IMSS sheared DNA library
JOURNAL
COMMENT      Unpublished (2000)
                Contact: Brendan J Loftus
                Department of Eukaryotic Genomics
                The Institute for Genomic Research
                9712 Medical Center Dr., Rockville, MD 20850, USA
                Tel: 301 838 0200
                Fax: 301 838 0208
                Email: enta@tigr.org
                Clones are derived from the Entamoeba histolytica HMI:IMSS sheared
                DNA library
                Seq primer: M13-Forward
                Class: Shotgun
                High quality sequence start: 86
                High quality sequence stop: 556.
                Location/Qualifiers
                    1. 1015
                        /organism="Entamoeba histolytica"
                        /strain="HMI:IMSS"
                        /db_xref="taxon:5759"
                        /clone_lib="Entamoeba histolytica Sheared DNA"
                        /note="Vector: pHOS1; Site.1: Bst I; Constructed at The
                        Institute for Genomic Research (TIGR), Rockville, MD.
                        Genomic DNA isolated from broth cultures of E. histolytica
                        using a method described by Clark and Diamond (Clark, A.
                        C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a
                        method for isolate identification. Exp. Parasitol.
                        77:450.). The DNA was mechanically sheared to give a
                        tight size distribution (-2 kb). The v + i method used for
                        the library construction is described in detail in Smith,
                        H.O. and Venter, J.C. (Making small insert libraries for
                        whole genome shotgun sequencing projects. In Genome
                        Sequencing: A Practical Approach, eds. M. Vaudin and B.

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 20:23:51 ; Search time 172 Seconds
(without alignments)
184.541 Million cell updates/sec

Title: US-09-391-606-7
Perfect score: 1166
Sequence: 1 MTKKHVAVVVEGILNRLPKQ.....DKLGSDPTFRKFDLGIISAF 217

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL17.*

- 1: sp.archaea.*
- 2: sp.bacteria.*
- 3: sp.fungi.*
- 4: sp.human.*
- 5: sp.invertebrate.*
- 6: sp.mammal.*
- 7: sp.mhc.*
- 8: sp.organelle.*
- 9: sp.phage.*
- 10: sp.plant.*
- 11: sp.podent.*
- 12: sp.virus.*
- 13: sp.vertibrate.*
- 14: sp.unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1162	99.7	422	2 Q927H6	Q927H6 chlamydia p
2	1162	99.7	422	2 Q9J5B3	Q9J5B3 chlamydia p
3	1156	99.1	715	2 Q46166	Q46166 chlamydia p
4	778	66.7	432	2 Q84628	Q84628 chlamydia t
5	667	57.2	207	2 Q46380	Q46380 chlamydia t
6	92.5	7.9	901	12 Q9Q3F9	Q9Q3F9 bovine aden
7	88.5	7.6	830	12 Q84501	Q84501 paramecium
8	88.5	7.6	902	12 Q912A7	Q912A7 bovine aden
9	85.5	7.3	684	2 Q9L1F8	Q9L1F8 streptomyce
10	85.5	7.3	836	12 Q9WAX4	Q9WAX4 chlorodella v
11	84	7.2	453	5 Q44021	Q44021 plasmodium
12	83.5	7.2	246	2 Q9L121	Q9L121 streptomyce
13	83	7.1	349	12 Q87696	Q87696 chimpanzee
14	83	7.1	743	1 Q58900	Q58900 pyrococcus
15	83	7.1	2233	2 Q9KG00	Q9KG00 streptococc
16	82.5	7.1	568	2 Q915N0	Q915N0 pseudomonas
17	82	7.0	1176	12 Q98587	Q98587 paramecium
18	81.5	7.0	395	5 Q26913	Q26913 trypanosoma
19	81.5	7.0	449	5 Q00819	Q00819 trypanosoma

20	81.5	7.0	900	12 Q91ZAB	Q91ZAB bovine aden
21	81	6.9	1128	2 Q9FDC0	Q9FDC0 bacillus th
22	80.5	6.9	214	2 P72800	P72800 synechocyst
23	80.5	6.9	409	2 Q68864	Q68864 aquifex aco
24	80.5	6.9	447	2 Q9AGU3	Q9AGU3 pseudomonas
25	80.5	6.9	469	2 Q50499	Q50499 streptomyce
26	80.5	6.9	902	12 Q9Q3G0	Q9Q3G0 bovine aden
27	80	6.9	644	1 Q9V057	Q9V057 pyrococcus
28	79.5	6.8	290	2 Q59078	Q59078 alteromonas
29	79.5	6.8	910	12 Q711A9	Q711A9 bovine aden
30	79	6.8	195	2 Q9KLF5	Q9KLF5 vibrio chol
31	79	6.8	473	2 Q9RB20	Q9RB20 erwinia car
32	79	6.8	482	10 Q80804	Q80804 arabidopsis
33	79	6.8	499	2 Q9ZL55	Q9ZL55 helicobacte
34	79	6.8	648	5 Q9VKU7	Q9VKU7 drosophila
35	79	6.8	650	5 Q9ULJ7	Q9ULJ7 drosophila
36	78.5	6.7	387	5 Q9VKT8	Q9VKT8 drosophila
37	78.5	6.7	412	2 Q50371	Q50371 methylophil
38	78.5	6.7	658	10 Q9FTM2	Q9FTM2 oryza sativ
39	78.5	6.7	3796	6 Q97699	Q97699 bos taurus
40	78.5	6.7	3796	6 Q9TTK4	Q9TTK4 bos taurus
41	78	6.7	753	10 Q9SUA4	Q9SUA4 arabidopsis
42	78	6.7	925	1 Q9HLE8	Q9HLE8 thermoplasma
43	77.5	6.6	696	8 Q9TIT2	Q9TIT2 phacelia ra
44	77.5	6.6	3186	4 Q9HI33	Q9HI33 homo sapien
45	77	6.6	505	1 Q28945	Q28945 archaeoglob

ALIGNMENTS

RESULT	1
Q927H6	
ID	Q927H6 PRELIMINARY; PRT; 422 AA.
AC	Q927H6
DT	01-MAY-1999 (TRENBLrel. 10, Created)
DT	01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT	01-JUN-2001 (TRENBLrel. 17, Last annotation update)
DE	CHLPN 76 KDA HOMOLOG_2 (CT623).
GN	CPN0729 OR CP0017.
OS	Chlamydia pneumoniae (Chlamydia pneumoniae).
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX	NCBI_TaxID=83558;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=CWL029;
RX	MEDLINE=99206606; PubMed=10192388;
RA	Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA	Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT	"Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL	Nat. Genet. 21:385-389(1999).
RN	[2]
RP	SEQUENCE FROM N.A.
RC	STRAIN=AR39;
RX	MEDLINE=20150255; PubMed=10684935;
RA	Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA	White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
RA	Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
RA	Bowman C., Dodson R., Gwinn M., Nelson W., DeBoy R., Kolonay J.,
RA	McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
RT	"Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT	pneumoniae AR39.";
RL	Nucleic Acids Res. 28:1397-1406(2000).
DR	EMBL; AE001654; AAD18868.1; -
DR	EMBL; AE002165; AAF37913.1; -
DR	TIGR; CP0017; -
KW	Hypothetical protein; Complete proteome.
SQ	SEQUENCE 422 AA; 48374 MW; ICFESB64B0ED3084 CRC64;

Query Match 99.7%; Score 1162; DB 2; Length 422;
Best Local Similarity 99.5%; Pred. No. 6.7e-99;
Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia. NCBI_TaxID=83558;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94156481; PubMed=7509320;
 RX Perez-Melgosa M., Kuo C.C., Campbell L.;
 RA "Isolation and characterization of a gene encoding a Chlamydia
 RT pneumoniae 76-kilodalton protein containing a species-specific
 RT epitope.";
 RL Infect. Immun. 62:880-886(1994).
 DR EMBL; L23921; AAA32117.1;
 SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 99.1%; Score 1156; DB 2; Length 715;
 Best Local Similarity 99.5%; Pred. No. 4.9e-98;
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCVWQWLV 60
 Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCVWQWLV 265

QY 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTTLNKENLAWFIGTGLGRLKAGDWSA 120
 Db 266 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTTLNKENLAWFIGTGLGRLKAGDWSA 325

QY 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180
 Db 326 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 385

QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217
 Db 386 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 2
 Q9JSB3 PRELIMINARY; PRT; 422 AA.
 AC Q9JSB3;
 DT 01-OCT-2000 (TremBLrel. 15, Created)
 DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
 DT 01-OCT-2000 (TremBLrel. 15, Last annotation update)
 DE CHLPN 76 KDA HOMOLOG_2 (C6233).
 GN CPJ0729.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10871362;
 RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
 RT from Japan and CWL029 from USA."
 RL Nucleic Acids Res. 28:2311-2314(2000).
 DR EMBL; AF002547; BAA9936.1;
 SQ SEQUENCE 422 AA; 48373 MW; FDC2F564B2679C0E CRC64;

Query Match 99.7%; Score 1162; DB 2; Length 422;
 Best Local Similarity 99.5%; Pred. No. 6.7e-99;
 Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCVWQWLV 60
 Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCVWQWLV 265

QY 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTTLNKENLAWFIGTGLGRLKAGDWSA 120
 Db 266 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTTLNKENLAWFIGTGLGRLKAGDWSA 325

QY 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180
 Db 326 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 385

QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217
 Db 386 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 3
 Q46166 PRELIMINARY; PRT; 715 AA.
 AC Q46166;
 DT 01-NOV-1996 (TremBLrel. 01, Created)
 DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)
 DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
 DE 76 KDA PROTEIN.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).

OC Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia. NCBI_TaxID=83558;
 OX [1]
 RN SEQUENCE FROM N.A.
 RP MEDLINE=94156481; PubMed=7509320;
 RX Perez-Melgosa M., Kuo C.C., Campbell L.;
 RA "Isolation and characterization of a gene encoding a Chlamydia
 RT pneumoniae 76-kilodalton protein containing a species-specific
 RT epitope.";
 RL Infect. Immun. 62:880-886(1994).
 DR EMBL; L23921; AAA32117.1;
 SQ SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;

Query Match 99.1%; Score 1156; DB 2; Length 715;
 Best Local Similarity 99.5%; Pred. No. 4.9e-98;
 Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCVWQWLV 60
 Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCVWQWLV 60

QY 61 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTTLNKENLAWFIGTGLGRLKAGDWSA 120
 Db 266 GKHSQVPWINGQKKPLLYLYGAFNLNPLAKATKTTLNKENLAWFIGTGLGRLKAGDWSA 120

QY 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180
 Db 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAAANYDPKEANSFTNYKGFSAIYMYGI 180

QY 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216
 Db 181 TDSLSFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216

RESULT 4
 O84628 PRELIMINARY; PRT; 432 AA.
 AC O84628;
 DT 01-NOV-1998 (TremBLrel. 08, Created)
 DT 01-MAY-2000 (TremBLrel. 13, Last sequence update)
 DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
 DE CHLPN 76 KDA HOMOLOG.
 GN CT623.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/OW-3/CX;
 RX MEDLINE=9900809; PubMed=9784136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RT "Genome sequence of an obligate intracellular pathogen of humans:
 RT Chlamydia trachomatis.";
 RL Science 282:754-759(1998).
 DR EMBL; AE001333; AAC68227.2;
 KW Complete proteome.
 SQ SEQUENCE 432 AA; 48365 MW; 46B3BA0A193A86BA CRC64;

Query Match 66.7%; Score 778; DB 2; Length 432;
 Best Local Similarity 63.8%; Pred. No. 1.5e-63;
 Matches 148; Conservative 26; Mismatches 42; Indels 15; Gaps 4;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDNNTFVPSSTTTEKAATNAMKYKCV 55
 Db 202 MAEKEYAVWVEAILNKLKPNFVVKVSVVDNNTLAKTNDPADASAAQPAKPT-KYDYL 260

QY 56 WQWLVGKHSQVPWINGQKKPLLYLYGAFNLNPLAK-----ATKTT--LNGKENLAWFI 105
 Db 261 WQWLVGKSTAMPWFNGQTKNLYGAYLEFNPLAEIPENKQSTTPTTKINGKENHAWFI 320

DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE HEXON (FRAGMENT).
OS bovine adenovirus 8.
OC Viruses; dsDNA viruses, no RNA stage; Adenoviridae; Adenovirus.
OX NCBI_TaxID=120509;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MSK/67;
RA Lemkuhl H.D., Hobbs L.A.;
RT "Partial sequence of the bovine adenovirus type 8 hexon gene."
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF238233; AAF63489.1; -.
DR InterPro: IPR000736; Adeno_hexon.
DR Pfam; PF01065; Adeno_hexon; 1.
DR ProDom; PD002815; Adeno_hexon; 1.
DR NON_TER 1
FT NON_TER 902 902
SQ SEQUENCE 902 AA; 102443 MW; 450C5E09AD3A2243 CRC64;

Query Match 7.6%; Score 88.5; DB 12; Length 902;
Best Local Similarity 20.7%; Pred. No. 12;
Matches 39; Conservative 29; Mismatches 77; Indels 43; Gaps 7;
QY 34 VPSETSTTEKAATNAMYKYCWQWLVGKHSQVPWINGOKKPLLYLYGAFLLNPLAKATKT 93
DB 603 IPNFTSTVVVNDPSWAFRCWSNRYKATETPMIGATKDPNLYSGSI--PYLDGT-- 658
QY 94 TLNGKENLAWFTGGLGRKAGDSATVRYEYVEALSVP-----EIDVSG--IGR 142
DB 659 -----FYLTHTF--QRVSIQWSSVPWPGDRLIPNWFETKRDYSVDAEGYNMSQ 707
QY 143 GNLLKFWFAQAATAANDPKANSFTNYKGFSAIYMGITDLSL-----FRAY 189
DB 708 CNITKDWYMIQMAYN---QAYOGYKLPVSHKYGFLENFQPMRSQVPIYGNLDFOLY 763
QY 190 GAYSPAN 197
DB 764 SAYIKSPN 771

RESULT 9
Q9L1F8 PRELIMINARY; PRT; 684 AA.
AC Q9L1F8;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE PUTATIVE NEUTRAL ZINC METALLOPROTEASE.
GN SC3D11.04C.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycineae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Brown S.P., Harris D.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RA Bentley S.D., Parkhill J., Barrell B.G., Rastam M.A.;
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
[3]
RP SEQUENCE FROM N.A.
RC STRAIN=A3(2);
RX MEDLINE=97000351; PubMed=8843436;
RA Redenbach M., Kieser H.M., Denapalte D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for

RT the 8 Mb Streptomyces coelicolor A3(2) chromosome.";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL157916; CAB76001.1; -.
DR InterPro: IPR001570; Peptidase_M4.
DR InterPro: IPR000130; Zn_MTPeptide.
DR Pfam; PF01447; Peptidase_M4; 1.
DR PRINTS; PR00730; THERMOLYSIN.
DR PROSITE; PS00142; ZINC_PROTEASE; UNKNOWN_1.
KW Protease; Metalloprotease.
SQ SEQUENCE 684 AA; 71289 MW; 594F70DDAB732BCF CRC64;

Query Match 7.3%; Score 85.5; DB 2; Length 684;
Best Local Similarity 23.8%; Pred. No. 16;
Matches 43; Conservative 23; Mismatches 62; Indels 53; Gaps 10;
QY 55 VWQWLVGKHSQVPWINGOKKPL-ILYGAFLNPLAKATKTTLNGKENLAWFIGTGLG-- 111
DB 417 VGDYLIGEEID---INGDTPLRYM-----DKPSKDGASKDSWYSG--IGVD 459
QY 112 -----LRKAGDSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANY 158
DB 460 VHYSSGANHFFYLLSESGTGKTINGVYNSPTSDGLPVTGIGRAKAKEIWF-RALTTFK 518
QY 159 DPKEANSTNYK-----FSALYMGITDLSLSPRA-----YCAYSKPANDKLGSDFT 205
DB 519 -----TSTNTYAGARTGTLLAAAGELYG-TDSAETAVANAWAGINVGARPGGDPDGT 572
QY 206 F 206
DB 573 F 573

RESULT 10
Q9WAX4 PRELIMINARY; PRT; 836 AA.
ID Q9WAX4
AC Q9WAX4;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
DE VCHTI-1.
GN VCHTI-1.
OS Chlorella virus.
OC Viruses; dsDNA viruses, no RNA stage; Phycodnaviridae; Phycodnavirus.
OX NCBI_TaxID=10507;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CVK2;
RA Hiramatsu S., Yamada T.;
RT "Chlorella virus vChTI-1 gene, complete cds."
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB022343; BAA78554.1; -.
DR HSP; P07986; 1EXG.
DR InterPro: IPR002965; P-rich_extensn.
DR PRINTS; PR01217; PRICHEXTENS.
SQ SEQUENCE 836 AA; 91356 MW; 6C283A494BA908E2 CRC64;

Query Match 7.3%; Score 85.5; DB 12; Length 836;
Best Local Similarity 21.8%; Pred. No. 21;
Matches 37; Conservative 20; Mismatches 44; Indels 69; Gaps 6;
QY 4 KHYAVVVEGILNR-----LPKOFFVKCSVVVDWNTVPSETSTT-----EKAATNAMY 51
DB 50 ENFTWFSEGLVRKGNKVMTIPK-----DWNKSIPTAGTTKIIPFGGVKALPCNLKY 100
QY 52 KCVQWLVGKHSQVPWINGOKKPLLYLYGAFLLNPLAKATKTTLNGKENLAWFIGTGLG 111
DB 101 NQIL--PLVGK-----DP-----S 112
QY 112 LRKAGDSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPK 161
DB 113 LAKRGKWTSKAVAPYVDACAFPTPDLPALSKASGLKFFTLAFTADSNK 162

Best Local Similarity	20.9%;	Pred. No. 11;
Matches	46;	Conservative
	26;	Mismatches
	82;	Indels
	66;	Gaps

SEQUENCE FROM N.A.
RC STRAIN=A3(2);
Thomson N.R., Parkhill J., Barrell B.G., Rajandream M.A.;
RA

QY 4 KHYAVVEGILNRL---PKOFFVKCSVVDNFTVPSTSTTEKAATNAMKYKCVQWLV 60
 DB 124 KHY-W--DAIRFYCAPPGVALLRCDNTNSGFMPCSKVAVSSCTRTMETQTPWFGFN 180
 QY 61 GKHSQ---VPWINGQKPLY---LYGAFLNPLAKATKTL-----NGK 98
 DB 181 GTRARENTHYWHGDRNPFLSNQYVNLNMCRRPNKVTPLVTLISGWFHSLPINDS 240
 QY 99 ENLAWFTGGTGLGURKAGW-----SATVRY-EVEALSVEIDVSGIGRG----- 143
 DB 241 PNQAWCFG-----GKWKDAIKEYQALVHPRTYGANNTDEINLTPPGGDPPEVTF 292
 QY 144 -----NLLKFWFAQAIAAAYDPKEANSFTNY 169
 DB 293 MWTNCRGEFLYCKMWNFLNVEDRNTANQPKHEHK-RNY 331

RESULT 14
 ID O58900 PRELIMINARY; PRT; 743 AA.
 AC O58900;
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-AUG-1998 (Tremblrel. 07, Last sequence update)
 DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
 DE HYPOTHETICAL 81.5 KDA PROTEIN PH1196.
 GN PH1196.
 OS Pyrococcus horikoshii.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=53953;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OT3;
 RX MEDLINE=98344137; PubMed=9679194;
 RA Kawaiaboyasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
 RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hosoyama A., Nagai Y.,
 RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
 RA Funahashi T., Tanaka T., Kudo Y., Yamazaki J., Kishida N., Oguchi A.,
 RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
 RA Masuchi Y., Shizuya H., Kikuchi H.,
 RT "Complete sequence and gene organization of the genome of a hyper-
 thermophilic archaeobacterium, Pyrococcus horikoshii OT3."
 RL DNA Res. 5:55-76(1998).
 KW EMBL; AP000005; BAA30296.1;
 DR Hypothetical protein; Complete proteome.
 SQ SEQUENCE 743 AA; 81475 MW; 6FEAF83BC01F3673 CRC64;

Query Match 7.1%; Score 83; DB 1; Length 743;
 Best Local Similarity 18.5%; Pred. No. 30;
 Matches 42; Conservative 35; Mismatches 76; Indels 74; Gaps 9;
 QY 9 VVEGILNRLPKOFFVKCSVVDNFTVPSTSTTEKAATNAMKYKCVQWLVGKHSQV 68
 DB 314 IIEELFNAPKOFIEVAEVSFYREI-----YYI-----NNGQW 346
 QY 69 INGKPKLYLYGAFLNPLAKATKTLNGKENLAWFTGGTGLGURKAGDWSATVRYEYVE 128
 DB 347 IDHKKEP-----SAKYITLSYRHIG-----GGVVVIGGFAGSLPADYKD 387
 QY 129 A-----LSVPEIDVSGIGRGNLLKFW-----FAQAIAAYDPKEANSFTN 168
 DB 388 APEAAKFTAFQTKVSGASSNLRLWLSAPLHMLNATVAYRFYVQANPGYDFPSPSIIDE 447
 QY 169 YKGFSAIYMGITDLSLFRAYGAYSKPAN--DKLSGDDTFERKFDLGI 213
 DB 448 ----ARIVVIRITDKGTWEL--AYALPSGVNLTGLTGTCKRLDASV 488

RESULT 15
 ID O9KGU0 PRELIMINARY; PRT; 2233 AA.
 AC O9KGU0;
 DT 01-OCT-2000 (Tremblrel. 15, Created)

DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-JUN-2001 (Tremblrel. 17, Last annotation update)
 DE BETA-GALACTOSIDASE PRECURSOR.
 GN BGAA.
 OS Streptococcus pneumoniae.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
 OC Streptococcus.
 OX NCBI_TaxID=1313;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=KMR7/87;
 RA Zaehner D., Hakenbeck R.;
 RT "The Streptococcus pneumoniae beta-galactosidase is a surface
 protein."
 RL J. Bacteriol. 0:0-0(2000).
 DR EMBL; AF282987; AAF97242.1;
 DR InterPro; IPR001649; Glyco_hydro_2.
 DR InterPro; IPR001899; Gram_pos_anchor.
 DR Pfam; PF00746; Gram_pos_anchor; 1.
 DR PRINTS; PR00132; GLHYDLASE2.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 DR PROSITE; PS00343; GRAM_POS_ANCHORING; UNKNOWN_1.
 SQ SEQUENCE 2233 AA; 247290 MW; 8170D06584D0A137 CRC64;
 Query Match 7.1%; Score 83; DB 2; Length 2233;
 Best Local Similarity 23.4%; Pred. No. 1.3e+02;
 Matches 57; Conservative 28; Mismatches 81; Indels 78; Gaps 14;
 QY 32 TFVPSSTSTTEKAATNAMKYKCVQWLVGKHSQVWINGOKKPLYLYGAF-----MNPL 87
 DB 483 THNPASEQTQIAAELGL-----LVOEEAFDTWYGG-KRP-YDYGRRFEKDATHP 531
 QY 88 AKA-----TKTTL-NGKENLA---WFIGGTLGLRKGADWSATVRYEYVEALSVEI 135
 DB 532 ARKGEKSDPDLRTMVERGKNPPAFMWSIGNEIGANGDAHSLATVK-RLVKV--IKDV 588
 QY 136 DVSGIGRGNLLKFWFA-----QAIAAYDPKEANSFTNYKGFSAIY-----MY 178
 DB 589 DKTRYVTMGADKFRFGNGSGGHEKIADELDAVGFN-----SEDNYKALRAKHPKLIY 642
 QY 179 GITDLSLFRAYGAYSKP-----ANDKLG-----SDTFERKFDLGI 213
 DB 643 GSETSSATRTGRSYRPERELKHSNGPERNEVEQSDYGNDRVWGKGTATATSWTFDRNAGY 702
 QY 214 ISAF 217
 DB 703 AGQF 706

Search completed: February 7, 2002, 21:41:09
 Job time: 4638 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:30:32 ; Search time 12230.8 Seconds
(without alignments)
1873.512 Million cell updates/sec

Title: US-09-391-606-4
Perfect score: 1389
Sequence: 1 atgttgtaactctattgtcc.....aagttctatgatccgagct 1389

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 segs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_em.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vi.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_on.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vi.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	1379.4	99.3	2238	6	AX045137	AX045137 Sequence
2	1369.4	98.6	2148	1	CHT76KDA	L3921 Chlamydia p
3	1359.4	97.9	2156	6	AX045131	AX045131 Sequence
c 4	1359.4	97.9	10421	1	AE001654	AE001654 Chlamydia
5	1359.4	97.9	12548	1	AE002165	AE002165 Chlamydia
c 6	1359.4	97.9	300550	1	AP002547	AP002547 Chlamydia
7	1347	97.0	1456	6	AX045135	AX045135 Sequence
8	1155.4	83.2	1852	6	AX045133	AX045133 Sequence
9	146.8	10.6	150	6	ARI44067	ARI44067 Sequence
c 10	56.6	4.1	10954	1	AE001333	AE001333 Chlamydia
c 11	54.4	3.9	7218	6	I66494	I66494 Sequence 14
12	53.6	3.9	1537	6	AX155814	AX155814 Sequence
13	51	3.7	5098	1	MCU51235	US1235 Mycoplasma
14	50.6	3.6	2085	6	AR008322	AR008322 Sequence
15	50.6	3.6	2085	6	AR027306	AR027306 Sequence
16	50.6	3.6	2085	6	AR035771	AR035771 Sequence
17	50.6	3.6	2085	6	AR037631	AR037631 Sequence
18	50.6	3.6	2085	6	AR078816	AR078816 Sequence
19	50.6	3.6	2085	6	AR079063	AR079063 Sequence
20	50.6	3.6	2085	6	AR084296	AR084296 Sequence
21	50.6	3.6	2085	6	I16740	I16740 Sequence 1
22	50.6	3.6	2085	6	I70472	I70472 Sequence 1
23	50.6	3.6	2085	6	I92332	I92332 Sequence 1
24	50.6	3.6	2086	1	STRSURPROA	M74122 Streptococc
c 25	50.2	3.6	14698	1	AE002357	AE002357 Chlamydia
26	49	3.5	423	1	MC049	Z33043 M. capricolu
27	48.8	3.5	10607	1	AE006310	AE006310 Lactococc
28	45.8	3.3	719	1	AF244521	AF244521 Streptoco
29	45.8	3.3	151847	2	AC022792	AC022792 Homo sapi
c 30	45.6	3.3	3720	14	S76368	S76368 ORF 5' of E
c 31	45.6	3.3	43658	14	HSV3PRGEN	H86409 Herpesvirus
c 32	45.6	3.3	112930	14	HSGEND	X64346 Herpesvirus
33	44.8	3.2	110000	2	PFMAL13P2_0	AL049185 Plasmodi
c 34	44.8	3.2	168310	2	AC090810	AC090810 Homo sapi
c 35	44.6	3.2	3882	10	AF104261	AF104261 Mus muscu
36	44.6	3.2	177205	2	AC079217	AC079217 Mus muscu
37	44.6	3.2	178419	2	AC074040	AC074040 Mus muscu
38	44.4	3.2	633	9	HSU23862	U23862 Human clone
39	44.4	3.2	149592	2	AC073077	AC073077 Homo sapi
40	44.4	3.2	173184	2	AC013583	AC013583 Homo sapi
41	44.2	3.2	175895	9	AC064862	AC064862 Homo sapi
42	43.8	3.2	463	3	PFA332C	M69163 P. falciparu
43	43.8	3.2	16377	3	AF202180	AF202180 Plasmodi
c 44	43.8	3.2	78578	2	PFMAL13P9	AL096783 Plasmodi
45	43.6	3.1	6744	1	SSEPFAA	X71880 S.suis (189

ALIGNMENTS

RESULT 1	AX045137	Sequence 7 from Patent WO0066739.	DNA	PAT	24-NOV-2000
LOCUS	AX045137				
DEFINITION	AX045137				
ACCESSION	AX045137.1	GI:11343736			
VERSION					
KEYWORDS					
SOURCE					
ORGANISM					
REFERENCE					
AUTHORS					
TITLE					
JOURNAL					
FEATURES					
Source					

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

CDS	766...2238	/note="unnamed protein product"	
		/codon_start=1	
		/transl_table=11	
		/protein_id="CAC17330.1"	
		/db_xref="GI:11343737"	
		/translation="MVNPIGPGPIDETPTPPADLSAQGLEASAAANKSAPFAQRIAGAE AKPKESKIDSVVERWSILRSVNALMSLADKLGIASSNSSSTSRSDVDTTATATP PPFFDYKDYQAOATAYDTIFSTSLADIQAALVLODAVTNKIDTADDETAIAAEW ETKNADAVKVAQITELAYASDNQAILDSGLKLTSPDLLQALLQASVANNKAAELL KEMDNVPVPEKPTAIQASLVQDQATADPDKNDGNIRADYFAGQNASGAVENAKSN NSTSNIDSAKQAATAKTQIAEAQKFPDPILOEABQVMVIOAKEDLNKIKPADGSDV PNPQTTVGGSKQSSGISIRVSLMDLAENETASILMSGFRQIMHMFNPFENDSQAA QOELAAQARAAGKAGDSDSAAALADAKALEAALGKAGQOQOQILNALQIASAANVSA GVLPLQQLVIRARYQAYVEQKLISEEDLNSAVDHHHHH"	
BASE COUNT	689 a 459 c 497 g 593 t		
ORIGIN			
	Query Match	99.3%; Score 1379.4; DB 6; Length 2238;	
	Best Local Similarity	99.6%; Pred. No. 0;	
	Matches 1383; Conservative	0; Mismatches 6; Indels 0; Gaps 0;	
QY	1 atgggttaattcctattggtccaggctctatagacgaacacagacacacccctcccgagat	60	
Db	766 ATGGTTAATCTTATGGTCCAGGCTCTATAGACGAACACAGACGACACCTCCCGCAGAT	825	
QY	61 ctttctgctcaaggattgagcagcagtcagcaataaagtcggaagctcaagaata	120	
Db	826 CTTTCTGCTCAAGGATTGAGCGAGTGCAGCAATAAGATGCGGAAGCTCAAGAATA	885	
QY	121 gcaggtgcggaagctaaagcctaaagaattcgaacccgattctgtagacgctgagcgtc	180	
Db	886 GCAGGTGCGGAAGCTTAAGCTTAAGCAATCTAAGACCGATTCTGTAGACGCTGAGCATC	945	
QY	181 ttgcttctgagtgaaatgctctatagctgagtcgagcagtaagctgagtccttctagt	240	
Db	946 TTGCGTTTGCAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT	1005	
QY	241 aacagctgtctctactagcagatctgcagcgtggactcaacacagcagcagccact	300	
Db	1006 AACAGCTGCTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACGACAGCGACGACCT	1065	
QY	301 acgctctctcccccacgctctgtgattattagactcaagcgcaaacagcttaacgatact	360	
Db	1066 ACGCTCTCCACCCACGCTTTCATGATTATAAGACTCAAGCGCAACAGCTTACGATACT	1125	
QY	361 atctttacctcaacatcactagctgacatacaggtgtgtgtgagcctccaggatgct	420	
Db	1126 ATCTTTACCTCAACATCATCTAGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCT	1185	
QY	421 gtcactaataaagatacagcgtctactgtaggaaacccgcaatcgctgcggagtgg	480	
Db	1186 GTCACATAATAAAGATACACGGCTACTGATGAGGAACACCGCAATCGCTCGGAGTGG	1245	
QY	481 gaaactaagaatgccgatgcaattaaagtggcgcgcaaaattacagaattagcgaatat	540	
Db	1246 GAAACTAAGAATGCCGATGCAGTTAAAGTTGCGCGCAAAATACAGAATTACGGAATAT	1305	
QY	541 gcttcgatacaacagcagcttcttactcttttagttaaactgaactcttcgacacctta	600	
Db	1306 GCTTCGGATAAACCAAGCGATTCTTACTCTTTAGTAACTGACTTCCTTCGACCTCTTA	1365	
QY	601 cagactgctctctccaatctgtagcaacaataaagcagcagctgagcttcttaagag	660	
Db	1366 CAGGCTGCTCTTCTCCATCTGTAGCAACAATAACAAAGCAGCTGAGCTTCTTAAGAG	1425	
QY	661 atgcaagataaaccagtagtccaggggaaacgcctgcaattgctcaatcttttagttgat	720	
Db	1426 ATGCAAGATAAACCAGTAGTCCCGGAAACGCTGCAATTCCTCAATCTTTAGTTGAT	1485	
QY	721 cagacagatgctacagcacacagatagagaaagatggaatgcatggtgggatcatat	780	

RESULT 2

CHT76KDA	2148 bp	DNA	BCT	17-MAR-1994
LOCUS	Chlamydia pneumoniae	76 kDa protein gene, complete cds.		
DEFINITION	Chlamydia pneumoniae			
ACCESSION	L23921			
VERSION	L23921.1	GI:435961		
KEYWORDS	76 kDa protein.			
SOURCE	Chlamydia pneumoniae DNA.			
ORGANISM	Chlamydia pneumoniae			
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.			
AUTHORS	Perez-Melgosa, M., Kuo, C.-C. and Campbell, L.			
TITLE	Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kilodalton protein containing a species-specific epitope			
JOURNAL	Infect. Immun.	62, 880-886	(1994)	
MEDLINE	94156481			
FEATURES	Location/Qualifiers			
source	1..2148	/organism="Chlamydia pneumoniae"		
		/db_xref="taxon:83538"		
CDS	1..2148	/codon_start=1		
		/transl_table=11		
		/product="76 kDa protein"		
		/protein_id="AAA23117.1"		

Qy	1	atggttaatcctctattggtccaggtcctctatagacgaaacagaacgacacactcccgcagat	60
Db	101	ATGGTTAATCCTATTGGTCAGGCTCTATAGACGAAACAGACGACACTCCCGCAGAT	160
Qy	61	ctttctgctcaagattggagcgagtgcgagcaaataaagatgcggaagctcaagaata	120
Db	161	CTTCTGCTCAAGATTGGAGGCGAGTCAGCAAAATAAGAGTGCAGAACTCAAGAATA	220
Qy	121	gcagggtcggaagcctaagcctaagaatctaagaccgattctgtagagcgatggagcattc	180
Db	221	GCAGGTGCGGAAGCTAAGCCTAAGAAATCTAAGACCGATTCTGTAGCGATGGAGCATC	280
Qy	181	tgtcgcttcgagtgaatgctctcatgagctctggcgagataaagctgggtattgcttctagt	240
Db	281	TTGGTTCTCGAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT	340
Qy	241	aaacgtctgttcttactagcatctgcagacgtggaactcaacgacagcgaccgcacct	300
Db	341	AACAGCTCGTCTTCTACTAGCAGATTGCAGACGTGAGCTCAACGACAGCGACCGACCT	400
Qy	301	acgctcctcacccacgctctgatattataagactcaagcgcaaacagctctacgatact	360
Db	401	ACGCTCCTCCACCACGTTTGATGATTAAAGACTCAAGCGCAACAGCTTACGATACT	460
Qy	361	atctttacctcaacatcacctagctgacatacaggtctgttgtagcctccaggatgct	420
Db	461	ATCTTTACCTCAACATCACTAGCTGACATACAGCTGCTTTGGTGAGCGCTCCAGGATGCT	520
Qy	421	gtcactaataaaggatacagcgctactgattgagaaacccaatcctcgcgagtg	480
Db	521	GTCACTAATAAAGGATACAGCGGCTACTGTATGAGAAACCCGAATCGCTGCGAGTGG	580
Qy	481	gaaactaagaatccgcatgcaattaaagtgtggcgcgcaaatcacgaattacgcaaat	540
Db	581	GAAACTAAGAAATCCGATGCAGTTAAAGTTGGCGCGCAAAATTACAGAAATPAGCGAAATAT	640
Qy	541	gcttcggataaccaagcattcttgactctttaggtaaacctgactcttcgcacctctta	600
Db	641	GCTTCGGATAACCAAGCGATTCTTGACTCTTTAGTAAACTGACTCTCTTCGACCTCTTA	700
Qy	601	cagactgctcttccaatctgtagcaaacataaacaacgactgagcttcttaagag	660
Db	701	CAGGCTGCTCTTCTCCAATCTGTAGCAACAATAAACAAACGCTGAGCTCTTTAAAGAG	760
Qy	661	atgcaagataaccgactagttccagggaacgcctgcaattgtctcaattttagttgat	720
Db	761	ATGCAAGATAACCCAGTAGTCTCCAGGAAACGCGCTGCAATTGCTCAATCTTTAGTTGAT	820
Qy	721	cagacagatgctacagcgacacagatagaaaagtggaaatcgatggggatgcataat	780
Db	821	CAGACAGATGCTACAGCGACACACATAGAGAAAGATGAAATGCGATTAGGGATGCATAT	880
Qy	781	tttcgagacagaacgctagttggagctgtagaaaatgctaaatctaataacagtaaacg	840
Db	881	TTTCGAGACAGAACGCTAGTGGAGCTGTAGAAATGCTAAATCTTAATAACAGTATAAGC	940
Qy	841	aactagattcagctaaagcagcaatcgctactgcttaagacacaaaatgctgaagctcag	900

```
/product="CT619 hypothetical protein"
/protein_id="AAD18866.1"
/db_xref="GI:4377032"
/translaton="MSFTSQNRRLRTISGSTRRIKLDHKYSGFDPRSPVAINLEELN
SGIYALRLHMLQSENTAAALNPNTTFTPTSTWTDYKHSRPOASSRAPSSQPT
DIVSAALALVLVDGGLAELVASVTEIDGLASTISTYRQLMASYLGLTTLTAQOEK
VFSFSSVPSEKLLHVKOEAKAEIOAKOEKAVLEAKGVSTEEATEALIKYVPIY
AOPFEKEIEPLHLYRAKVCAPLOEMNENAIOLTPPAITPDNNEVANGMTLSTI
LOALDAIKAPALGGQDEIITILQTLPLVDKTTFKAEFDLLIYATQLPNTASLAL
YLDTRQAEIKTKIVYQNSISLSTKRVNRRNSMLETQLSMFOOACNCFVTWIS
QANALNATNKIYASVLTSMEMYGGLLSYMERLADDEKAIFDKSNEXYLPHI
VGSWYNGHIAKMAAYELAETSLGTAVTSDQIKAYLQTRGNEFKATRHFFHNGD
QMTQFNEFVGNCLTTANGAIOPDLGCFIREAMTVGTVEADYVNAQRIILNEFTA
ATRAHLOLQIAELOKKADDDPGKASFENRKEFAVAAMITSESLGDALISMLNSQ
LPKOEALFKPLIEINENNAALNLSLLOITNEFTSVYVYSLSSVLYOSKQGNLF
AGDYETFLAAAREREYIYDTRAKCAINLYNGLLOKINSLPGATSAQKQEMNL
YQYLSVTLNQLTVLESLLAGLRLQTTSSNNKYDKSVFKIESFDDWIPTLAALESF
LTSGFNFISATGLGLPLFTQVSDQQTYSQGTQQLNLQNMFTTIOQEWTLVSTMQ
VLNGLISQLAGAIYSN"
complement(3078..5033)
/gene="Cp00728"
CDS
complement(3078..5033)
/gene="Cp00728"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_1 (CT622)"
/protein_id="AAD18867.1"
/db_xref="GI:4377033"
/translaton="MVPVIGPGPIDETERTPPADLSAQGLEASAANKSABEAQRIAGE
AKPKESKTSVERMSILSRNALMSLADKLGIASSNSSSTSRSDVDSSTTATPTP
PPTFDYKTAQAYDTIFTSTLADIAQLVSLQDAVTNKIDTATDEETAIAAEW
ETKNADAVQAQGTTELAKYASDQNALILDSGLKLTSPDLQALQLQSVANNAAEEL
KEMDNPVPGKTPAIAQSLVDOTDATOI EKDGNAIRDAYFAGONASAVENAKSN
NSLSDSAKATATAKTOIAEAQKFPDPSPILOEAEQVIAEKDLKNIKPADGSDV
PNRGTTVGGSKQSGSGSIRVSMILDOAENETILMSGRFQRMHMFENTPDQRA
QOEIAQARAAGDQSAALADAKALEALGKAGQOQGLNALGQIASAAVSA
GVPPAAASSIGSSVKYKTSKSTGSDYKTOISAGYDAYKSINDAYGRARNDRDVI
NYSVTPALTRVSPRARTARGPKTDQALRIVSGNSRTGLDVIYQVSAQSQWQITQ
SNPOANNEIRKLTSAVTKPPQFGYPIYVOLSNDSTQKFIKLESPLFAEGSRTAEEK
ALSPETNSLFIQVVLNIGSLYSGYLQ"
complement(5145..6413)
/gene="Cp00729"
CDS
complement(5145..6413)
/gene="Cp00729"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_2 (CT623)"
/protein_id="AAD18868.1"
/db_xref="GI:4377034"
/translaton="MKKQVYQWLASVLLALTISGYAELPLSEQVKSKSHYTTILDEVK
DYLKRGFVETRQDQVLRAGDVRARMLYFREDIKNPSDKDKYNPLPNVRYSSEYVL
YIDYRAERNLWSKMNWTAAGGENTAAGVDINRAFLGYRFYKNPRTDFFMEIGRS
GLGLDFSEVQFQSNFDGLHIYMTRELSKDYPIQVIVHGGPFVVMNTKHYANVVEGI
LNRLPKQFFVKCSVVDNTEVPSETSTTEKAAANAMKYKVCVQWLGVKHSQVPMWG
KQRPYLDIGAFMLNPLAKATKTLNGKENLAWFIFGGTLGLGRKAGDWSATVREYVEA
LSVPEIDVSGIGRNLKFWFAQIAAANYDPKPEANGFTNYKGSALYMYGIDLSLSPR
AYGASXPANDKLGSDFTFRKFDLGIISAF"
complement(6516..8159)
/gene="mvn"
CDS
complement(6516..8159)
/gene="mvn"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAD18869.1"
/db_xref="GI:4377035"
/translaton="MSRKDNVSLARSIFNLSTGFCSTRITGIFREIATAMATYFADPI
VFAAFGLFRVFFVRLKGLLILEQAFIPHFELRQSLDRAAFFRRFRSLKLGSTI
IFTLLEAVLVLYQVVEEGTYDMLTLLMPCGIFLMYNNVNGALLHCENFFGVG
LAPVNNIIMFVFIARHSDPREIRIIGLSVALVIGFFFWLITVPQVWVLEAKSP
POEHSVRALLALSLGILTSFQNLNLSDICLARVHEIGPIYLMYSIKIYOLPHI
LFGFVGFVTLVLLPAISRCVQRHEDHERGLKMKFVLTLTMSVYMIITAGLLLLALPGVRV
```

```
LYEHGLFPOSAYVAIVRLRGYCASIIIPMALAPLVLVFAOQRYAVPFLIGITALA
NIVLSVLGRWLKDSGISATSIATWOLYIFLWYSSRRRLPMYSKLLMESIRRSIK
VMGTMLACMITLGLNLTQTYYIFLNPITPLAWPLSSITAQAIAPLSESCIFLAPL
FGFAKLLRVDELINLASFEYWRGQGLLQRHVMQDTON"
8143..8319
/gene="Cp00731"
CDS
8143..8319
/gene="Cp00731"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAD18870.1"
/db_xref="GI:4377036"
/translaton="MSFLLLKPFSPPLKODYLFDISPYTSSSETTIGGSYFKLNKASL
QSSTLRLSISISIS"
8654..9535
/gene="nfo"
CDS
8654..9535
/gene="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AAD18871.1"
/db_xref="GI:4377037"
/translaton="MKVLPSPISPLLGAHTSTAGGLKNAIYEGRDIGASTVOIFTANQ
RQWRALKEEVIEDFKALKETDLSYIMSHAGLYINPGADPVLVLEKSRIGIYQEIL
DCITLGSIFVNFHFGAALKSSDKMKIVSSFSQAPLDFSSPPLVLLLETTAGQGT
LIGSFEELGYLVONLKNQIPIGVCDVTHIFAAQYDITSPQWEDYNLFEDYVGLS
LYRAFHLNDSMFFLQANKRHPALGEGYIGKESFKFLMTDERTRKIPYLETPGPN
WQKEIGLLKFKSKNRDS"
complement(9654..10283)
/gene="rs4"
CDS
complement(9654..10283)
/gene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AAD18872.1"
/db_xref="GI:4377038"
/translaton="MARYCQPKNRVARRFGANIFGRSRNPLLLKPPHPPGQHGMRKKK
SDYGLQLEERQKLKACYMIMERQVLFKAFKEVHKQNVQAMFELERFECRLDNNVYRM
GFARTIPAAQLVAHQHILVNRVDRRSPFLRPMQOISLKEKSKRQSVKDALESKD
ESSLPYSILDKTQKGEILLVSPEDQIEAQLPLPINISVVCFLSHRT"
BASE COUNT 3278 a 2158 c 1983 g 3002 t
ORIGIN
Query Match 97.9%; Score 1359.4; DB 1; Length 10421;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 1 atgtttaatctatgtgtccagggtctatagacgaacacagcacacccctccgcagat 60
Db 5033 ATGTTTAACTCTATTGTGTCAGGCTCTATAGCAAAACACACGACACCTCCCGCAGAT 4974
Qy 61 cttctctcctcaagattggagccagtcacgaataagagtcggaagctcaagaata 120
Db 4973 CTTTCTGCTCAAGGATTGGAGCGGAGTGCACCAATAAGATGCGGAAGCTCAAGAATA 4914
Qy 121 gcaggtgcggaagctaaagcctaaagaattcttaagaccgattctgtagacgattggagcattc 180
Db 4913 GCAGGTGCGGAAGCTAAGCCTAAGAAATCTAAGACCGATTCTGTAGACGATGAGCATC 4854
Qy 181 ttgcgtctgcagtgaaatgctctcatgagctgcagataagctgggtgattgcttctagt 240
Db 4853 TTGCGTTCTGCAGTGAATGCTCTCATGAGCTGTCAGATAAGACTGGGTGATTGCTTCTAGT 4794
Qy 241 aacagctcgtcttactagcagatcttcagacgtggactcaagacagcagccacact 300
Db 4793 AACAGCTCGTCTTCTACTAGCAGATCTCGACACCTGGGACTCAACGACAGCAGCCGACCT 4734
```

Db 3653 G 3653

RESULT 5

AE002165

LOCUS

DEFINITION

AE002165.1 12548 bp DNA BCT 30-MAY-2000

Chlamydomophila pneumoniae AR39, section 2 of 94 of the complete genome.

ACCESSION

AE002165 AE002161

VERSION

AE002165.1 GI:7188948

KEYWORDS

SOURCE

ORGANISM

Chlamydomophila pneumoniae AR39.

Chlamydomophila pneumoniae AR39.

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.

REFERENCE

AUTHORS

Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

TITLE

Genome sequences of Chlamydia trachomatis Mopn and Chlamydia pneumoniae AR39

JOURNAL

Nucleic Acids Res. 28 (6), 1397-1406 (2000)

MEDLINE

20150255

PUBMED

10684935

REFERENCE

AUTHORS

2 (bases 1 to 12548)

Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Winn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

TITLE

Direct Submision

JOURNAL

Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA

FEATURES

source

1. 12548

Location/Qualifiers

/organism="Chlamydomophila pneumoniae AR39"

/strain="AR39"

/db_xref="taxon:115711"

/note="synonym: Chlamydia pneumoniae AR39"

297. .965

/gene="CP0011"

297. .965

/gene="CP0011"

/note="similar to SP:P31218 GB:X71492 PID:296347 GB:U00096
PID:1736770; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="uridine kinase"

/protein_id="AAF37907.1"

/db_xref="GI:7188950"

/translation="MLMLMLIIIGTGGGAGKTTLTQNIKEIFGSDVSVCODNYIK
DRSHYPERANLIWDHPADFNDLLIDFKRLKNEIQAQPVDFVGNRKSRETEIK
IYPSKVLIGILVENEQLRDLMDIRFDVDPADREILRMVRDVOEGDSVDCIMIS
RYLSWAKPMKEKIEPTKRYADIIIVGNRQNVVTVLLSQIKNHLNLESDETIYN
VNSK"

complement(1047. .2021)

/gene="CP0012"

complement(1047. .2021)

/gene="CP0012"

/note="conserved hypothetical protein; identified by
Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF37908.1"

/db_xref="GI:7188951"

/translation="MEKKYVALAYVITRVDPNPEETLHKKTELDLDVSCRIVISEP
GINKQSGYEPHEALYMQWLKERNFSKIKFKHIIKENIFPRITVYKRELAAALGCE
VDLSQAKHISPOEHWELKQENRFLIDVRNNYENIGHDFDNATLSDIQITFRFFETPA
EKLAQECDFPTVMVYTCGIRCELYSLEKGFKEVYQLDGGVTAIQFGDSVGTGKIF
LGLIFVDDRLAIPIDESDPPVAFIAPCCCHQTPSDAYNACDGNALFLCQDEGIC


```

QHGCCGEGCSQSPVRKFDSDSRGNKPFRAHLCEISENSSESACCL1"
2238. .2867
/gene="CP0013"
CDS
2238. .2867
/gene="CP0013"
/note="similar to SP:P81288; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S4"
/protein_id="AAF37909.1"
/db_xref="GI:7188952"
/transl_table="MARIYCGPKNVRARFGANFGRSRNPLKKPHPGHGMQRKKK
SDVGLQEEQKQKACACYMIMKOLVKAKEVHKQCNVAQMFLERCECLDNNVYRM
GFAKTIFAAQLVAHQHILVNGRVRDRSFLEPGMOISLKEKSKRLQSVKDALESKD
ESSLPYSISLQKCGFKGELLVSPQDQIEAQLPLPINISVVCFLSRT"
complement(2986. .3867)
/gene="CP0014"
complement(2986. .3867)
/gene="CP0014"
PID:405898 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="endonuclease IV"
/protein_id="AAF37910.1"
/db_xref="GI:7188953"
/transl_table="MKVLPPTSLPGAHTSTAGGLKNAIYEGRDIGASTVOIFTAHQ
RQWRRAKKEVEDFKAALKEDLSTIMSHAGLILNPGAPDPVILEKSRIGIYEIL
DCITLGSFVNFHQAALKSKEDKMKIVSSFSQSAFLDSSPVLVLLETAGQGT
LISNFEELYLQNLKNOIPIGVCDTCHIFAAGYDITSPOGWEDVILNFEDEVGLS
YRAFHLNDSMFLGANKRHPALGEGYIGKESFKFLMDERTRKIPKYLETPGPGEN
WKEIGELLAFKSNRDS"
complement(4202. .4378)
/gene="CP0015"
complement(4202. .4378)
/gene="CP0015"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:7188954"
/transl_table="MSFLIKPFPPLKQDYLFDISPYTSSETTIGGSYFLNKASL
QSSTLRLRSIIS"
4362. .6005
/gene="CP0016"
4362. .6005
/gene="CP0016"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:7188955"
/transl_table="MSRKDNVLSRISFNLSTFCSTRITGIFREIATAMATFCADPI
VAFWLGFRTVFRLKTLGLLEQAFIPHEFLRAQSLDRRAFFFRFRSLIKGSTI
IFTLLEAVLWVQLVVEGTYDMILLTLLPCGIFLMNINVGALLHCENKFFGVG
LAPVNNIIFWVAARHSDPRRIIGLSVALVIGFFEWLITVPQWVKFLLEAKSP
POHDSVRALLAPLSGLISFISFQNLNDSICLIARYHVEIGPLYLMYSKRYQLPIH
LFGFVTVLPAISRCVQRDEHGLKMKFVLTMSVMIWTAGLLALLALPGVRV
LYEHLFPQSAVAIVRLVYGASIIIPMALPLVSVLFAQRQYVLPFLGIGTALA
NIVLSVLGRWLVKDVSGISYATSIWVQLYFLWYSSKRLPMYSKLLWESIRRSK
VMTGTMACMITGLNLIOTTTVIFLNPLTLPANPLUSSITAIQAFLSSECFIAPFL
FGAKTLRVEDLINLASFEYWRGQGLLQHQHMQDTQN"
6108. .7376
/gene="CP0017"
6108. .7376
/gene="CP0017"
/note="conserved hypothetical protein; identified by

```

```

Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:7188956"
/transl_table="MKQVYQWLASVLLALTISGYAELPLSEQVKSHYTTLLDEVK
DYLKRGFVETRKQDGLRIAGDVRABMLVFEREDIKNPSDKDKNPLPVARYSEFVL
YIDYRAERNLSSKMMWTATAGSENTAAGVDINRAFLGYRYKNPRTTRTFDFMEIGRS
GLDLFESEVQFSNFDGLHIYTRLSLQDPYQVIVHGGPFVYVMMTKKHAYAWVEGI
LNLKPOFFVYKSVYDNTFVPSSTTERKAATNAMKYKCYVMQWLKRSQVFWING
QKPLLYLGAFALMNLAKTTLTKNGENLAWFIGGTLGLLRKAGDMSATVRYEYEA
LSYPEIDVSGIGRNLAKLFWFAQAIAANYDPKEANGFTNYKGSALYMYGTTLSLSPR
AYGASKPANDKLGSDTFKFDLGIISA"
7488. .9443
/gene="CP0018"
7488. .9443
/gene="CP0018"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:7188957"
/transl_table="MVNPIGPGPIDETERTPPADLSAQGLEASAANKSAQAORTAGAE
AKPKESKTSVERWSILRSVNALMSLADKLGITASSNSSSSTSRSDVDTATATPP
PPFTDDYKQAOATYDTITFTSLADIQAALVSLQDRAVYNIKDITATDEETATAEAE
ETKNADAVKQAQITELAKYASQNDQAILDSGLKLTSLDQALLOALQSVANNKAAELL
KEMQDNVPVQKTPAIQAQSLVQDQATATQTEKDGNARIDAYFAGQNAAGAVENAKSN
NSINSDIAAATAATAKTAIAEKFPDPSPILOEAQPMVIAEKDLNKTKPADGSDV
PNPQTTVGGSKQCSSTIGSRVSLMDDAENETAILMSGFROWIMHMENTENPDQAA
QVELAQAARAKAGDSDAAALADAKALEALGKAGQOQILNALGOLASAAVVA
GVPPAAASSIGSSVKQLYKTSKTSQSYKYQISAGYDAYKSINDAYGRANDATDVTI
NNVSTPALRSVPARTAEKPEKTDQALARKVIGNSRTLGVDVTSVALQSVMSQVLIQ
SNPOANNEIRQKLTSAVTPPOFGYPYVOLSNDSTQKFTAKLESFAEGSRATAEIK
ALSFTNSLFIQOVLVNIIGSLYGLQ"
complement(9770. .12388)
/gene="CP0019"
complement(9770. .12388)
/gene="CP0019"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37915.1"
/db_xref="GI:7188958"
/transl_table="MSTFSIQNRLTISGSTRIRIKLDHRYSGFDRSPVAINLEELN
SGIYALRLNALQSENNTVAALINPNNTIPPTTSMTDYKHSRPOASSPRAPSOTPT
DIYSAALALVLDIGLAELVASVTEIDIGALSTISTVRLMASYLGLTLTAEQEK
VPSVSSVPSKELNLEHVKEAKAEIOAKOEIKAVLEAKGVSTEEETAILKEPYDIY
ADPFKEFIEEPLHYRKYKAPIQEMENAIQLLTPPALTPDNVNEVNGMNTLSLI
LQADDAIKQAPALGGQDEIITLIQTLVPLVDTFTFAEFDLIYATQLPNTASLKL
LYTDROIAYRGKITVYQNSIQNLSEKRVENNRSMLETSQMFQQAQCNCFVTHI
QANALNIAITNKYISAVLTTSMYEGGLCLISYMYERLADDEKAIKFKSVNEYLPFI
VVGSSVWNGWIAKMAAYOELAEYSLGTAVTSDQIKAYLOTGRNEKATRHEFINIGD
OMQVQNETVFGNLTTPANGAIQDPLGFFIREAMTVNGTVTEADYVSNQORILNFEIN
ATAHLOLQIQLAELOKKAADDLPFGKASFTENKRFVAAWITSESDALISMLNSQ
LPKQEAFLKPLIEINFNNLAALNSLLQITNEFTSVIYSLVSVQSVSKTGQNLF

```

Query Match Score 1359.4; DB 1; Length 12548;

Best Local Similarity 99.5%; Pred. No. 0;

Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 atgtgtaactctatgtccaggctctatagacgaacacagacacaccccccagat 60

|||||

Db 7488 ATGTTTAAATCCATTGTGTCAGGTCCTATACAGAAACAGACACACCTCCCGCAGAT 7547

|||||

Qy 61 cttctcgtcaagatggagcgagtgacgaataagagtcggaagctcaagaata 120

|||||

Db 7548 CTTTCTGCTCAAGATTGGAGGCGAGTGCACGAATAAGTAGTGCAGAGCTCAAGAATA 7607

|||||

Qy	121	gcagggtcggaagc	taagcctaaagaat	ctaagaaccgattct	gttagagcagatg	agacatc	180
Db	7608	GCAGGTGCGGAAGCT	TAAGCCTAAAGAACT	TAAGACCCGATCT	GTAGACGCGATG	GAGCAGATC	7667
Qy	181	ttcgcttctcagtg	aatgcctcatgagct	tgccagataagct	gggtattgctct	tagt	240
Db	7668	TTGGGTCTGCAATG	CTCATGAGTCT	GGCAGATTAAGCT	TGGGTATTCCTCT	TAGT	7727
Qy	241	aacagctcgctctt	ctactagcagatct	gcagacgtggact	caacgacagcgacc	gcacct	300
Db	7728	AACAGCTCGCTTCT	TACTAGCAGATCT	GCAGACGTGGACT	CAACGACAGCGAC	CGACCT	7787
Qy	301	acgcctctccacca	cgctctgatatt	ataagactcaagc	gcaaacagctt	acagatact	360
Db	7788	ACGCTCTCCACCC	ACGTTTGATGATTA	TAGATCTAAGCGCAAA	CGCTTAGCATACT		7847
Qy	361	atcttaccctcaac	atcactagctgac	atcacagctgct	ttggtgagctccag	gatgct	420
Db	7848	ATCTTTACCTCA	ACATCACTAGT	GCACATACAGGCT	GCTTTGGTGAGCTC	CCAGGATGCT	7907
Qy	421	gtcaactaataa	agagatacagcg	gctactgatgagga	aaacgcgaatcg	ctgcggagtgg	480
Db	7908	GTCACTAAATATA	AAAGGATACAGCG	CTACTGATGAGAAAC	CGCAATCGCTCGGAGTGG		7967
Qy	481	gaaactaagaat	gcggatgcaatt	aaagctggcgcgca	aaattacagaatt	tagcgaatat	540
Db	7968	GAACATAAGAT	ATCCGATCAAGT	TGCGCGCAAAAT	TACAGAATTAGCA	GAATAT	8027
Qy	541	gcttcggataac	caacgagctctt	gactcttagt	aaactgacttcc	ctcgaccttta	600
Db	8028	GCYTCGGATTA	CAACGCGATCT	CTGACTCTTTAGT	AAACTGACTTCT	CTCGACTCTTA	8087
Qy	601	cagactgctcttc	ccaatctgacaa	caataacaagcag	ctgagcttctt	aaagag	660
Db	8088	CAGCTGCTCTT	CTCCANCTGTAG	CAACAATACAAAG	CAGCTGAGCTTCT	TTAAGAG	8147
Qy	661	atcgagataac	cccgatgtccc	agggaaaaacgct	gcaattgtc	caatttttagttgat	720
Db	8148	ATGCAAGATTA	ACCAGCTAGT	TCCAGGAAAACG	CGCTCAATTTGCT	CAATCTTTAGTTGAT	8207
Qy	721	cagacagatct	cacgcacacag	atagagaagat	gaaatcgattg	gggagtgcatat	780
Db	8208	CAGACAGTCT	TACGCGACACAG	TAGAGAAGAT	GGAAATGCGAT	TAGGGATGTCATAT	8267
Qy	781	tttcaggacag	aacgctagtgag	ctgtagaaaat	gctaaatct	ataaacagataagc	840
Db	8268	TTTGCAGGAC	AAGCCTAGT	GGAGTGTAGAAAAT	GCTAAATCTAA	TAAACAGTATAAGC	8327
Qy	841	aacatagattc	agctaagcagca	atcgctactgt	ctagacacaaaa	tagctgaagctcag	900
Db	8328	AACATAGATT	CAGCTAAAGCAGCA	ATCGCTACTGT	TAGACACAAATAG	CTGAAGCTCAG	8387
Qy	901	aaaaagtccccc	gactctccaatt	cttcgaagc	ggaacaaaatg	taatacaggtctgag	960
Db	8388	AAAAAGTTCC	CCGACTCTCCAAT	TCTCAAGAAGCG	CAACAAATGGT	TAATACAGGCTCAG	8447
Qy	961	aaagatctt	aaaaatatca	aaactgcagatg	ttctgtatgtt	ccaaatccaggaactaca	1020
Db	8448	AAAGATCTT	AAAAATATCAAC	CTCGATGGTCT	GTATGCCAAATCC	AGGAACATACA	8507
Qy	1021	gttgaggtcc	caagcaacga	gaagtattgt	agtattgctgtt	ctcatctgtta	1080
Db	8508	GTTCGAGGCT	TCCAGCACAACGA	AGTAGTATTTGGT	TAGTATTCGT	TGTTTCCATGCTGTTA	8567
Qy	1081	gatgatgct	gaaatgagacc	gcttccatttt	gtatgtcgtgg	tttcgtcagatgattcac	1140
Db	8568	GATGATGCT	GAAAATGAGAC	CGCTTCCATTTT	GATGCTGTTT	TCCATGCTGTTA	8627
Qy	1141	atgttcaata	cggaaatcct	gatctcaagct	gcccaacagg	gagctcgacacagct	1200
Db	8628	ATGTTTCAAT	ACGGAAATCT	GATTCTCAAGT	CTGCCAACAGG	AGCTCGAGCACAGCT	8687
Qy	1201	agagcagc	gaagccgctg	agatdacat	gctgctgcag	cgtctgcaaa	1260

Db	8688	AGACGACGAAGCCGCTGGAGATGACAGTGTCTCTCAGCGCTGGCAGATGCTCAGAA	8747
Qy	1261	gctttagaagcgctctaggttaaagctgggcaacacagagcatactactcaatgctttggga	1320
Db	8748	GCCTTAGAAGCGGCTCTAGGTAAAGCTGGCGCAACACAGGGCATACTCAATGCTTTAGGA	8807
Qy	1321	caagtcgcttctgctgctgtgtgagcgaggag-tctctcccgctgcagcaagtctatg	1379
Db	8808	CAGATCGTCTCTGCTGTGTGTGTGAGCGCAGGAGTCTCTCCGCTGCAGCAAGTCTATA	8867
Qy	1380	g	1380
Db	8868	g	8868
RESULT	6		
AP002547/c			
LOCUS			
DEFINITION		AP002547 300550 bp DNA BCT 08-DEC-2000	
ACCESSION		Chlamydothila pneumoniae J138 genomic DNA, complete sequence,	
VERSION		section 3/4.	
SOURCE		AP002547 AB033786 AB033787 AB033816 AB033817 BA000008	
KEYWORDS		AP002547.2 GI:10176693	
ORGANISM		Chlamydothila pneumoniae J138 (strain:J138) DNA.	
REFERENCE		Chlamydothila pneumoniae J138	
AUTHORS		Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.	
TITLE		1 (sites)	
JOURNAL		Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K.,	
MEDLINE		Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.	
AUTHORS		Comparison of whole genome sequences of Chlamydia pneumoniae J138	
TITLE		from Japan and CWL029 from USA	
JOURNAL		Nucleic Acids Res. 28 (12), 2311-2314 (2000)	
MEDLINE		20330349	
AUTHORS		2 (bases 1 to 300550)	
TITLE		Shirai, M.	
JOURNAL		Direct Submission	
COMMENT		Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases.	
FEATURES		Mutsunori Shirai, Yamaguchi University School of Medicine,	
source		Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi	
		755-8505, Japan (E-mail:mshirai@po.cc.yamaguchi-u.ac.jp,	
		tel:81-836-22-2227, fax:81-836-22-2415)	
		On Sep 15, 2000 this sequence version replaced gi:6172398	
		gi:6172300 gi:6172396 gi:6172398 gi:8978889.	
		AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).	
		Location/Qualifiers	
		1. 300550	
		/organism="Chlamydothila pneumoniae J138"	
		/strain="J138"	
		/db_xref="taxon:138677"	
		/notes="synonym:Chlamydia pneumoniae (strain J138)"	
		complement(21..794)	
		/gene="dapF"	
		complement(21..794)	
		/gene="dapF"	
		/codon_start=1	
		/transl_table=11	
		/product="diaminopimelate epimerase"	
		/protein_id="BAA98725.1"	
		/db_xref="GI:8978890"	
		/translation="MAFYSPSTISKYFIYSGAGNRFLIGTLPEDVRFLOCOTRVDR	
		GFLYLPSSCADAQLIFNSDGRPTMCGNLCAHAHLASOKGSDISVSTOSGLYS	
		GYFYSDRWLVDMTLADSRVSRHRLSRPDLPEKVCVHTGPHAVVILPELISLVD	
		SILGPELRHQTSTPDGVNVNFVILGHCOLRKYTVRGEVGETAGCTGALASALV	
		SNSYGNKESIQTHTWGEELMTVSQNRVYLGVSVRDL"	
		complement(763..1338)	
		/gene="clpP_1"	
		complement(763..1338)	
		/gene="clpP_1"	
		/codon_start=1	
		/transl_table=11	
		/product="CLP protease"	
		/protein_id="BAA98726.1"	
gene			
CDS			

Query Match	97.9%	Score 1359.4	DB 1	Length 300550
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1374	Conservative	0	Mismatches	6
Indels	1	Gaps	1	

QY	1	atggttaactctattggtccaggtctctatagcaaaaacagacacacotcccgcagat	60
DB	217598	ATGGTTAACTCTATTGGTCCAGGTCTTATAGACGAAACAGAACACACCTCCGCAGAT	217539
QY	61	ctttctgctcaaggattggaggcgagtcgacaaataaagatcgcggaagctcaagaata	120
DB	217538	CTTTCTGCTCAAGGATTGGAGCGAGTCGAGCAATAAAGAGTCGGAGGCTCAAGAATA	217479
QY	121	gcagggtcggaagctaagcctaagaacttaagaccgattctgtagagcgatggagcatc	180
DB	217478	GCAGGTGCGGAAGCTAAGCCTAAGAACTAGACCCGATTCTGTAGACGATGGAGCATC	217419
QY	181	ttgcgttctgcagtgaaatgctctcatgagctggtggcagataagctgggtattgcttctagt	240
DB	217418	TTGCGTTCTCGATGAATGCTCTCATGAGCTTCGCGAGATAAGCTGGGTATTGCTTCTAGT	217355
QY	241	aacagctcgctcttctactagcagatctgcaacgtggactcaacgacgacgacgcacct	300
DB	217358	AACAGCTCGCTTCTTACTTAGCAGATCTGAGACGTGGAGCTCAACGACAGCAGCCGACCT	217299
QY	301	acgctctctcaccccgctctgatgattataagactcaagcgcaaacagcttaagcatact	360
DB	217298	ACGCTCTCTCCACCCAGTTTGATGATTATAAGACTAAGCGCAAAACGTTACGATACT	217239
QY	361	atctttacctcaacatcactagctgacatacagggctgtttggtgagctccagatgct	420
DB	217238	ATCTTTACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTGAGGCTCCAGGATGCT	217179
QY	421	gtcaactataaagatatacagcgctactgatgaggaacccgcaatcgctgcggagtg	480
DB	217178	GTCACTAATATTAAGGATACAGCGCTACTGATGAGAAACCGCAATCGCTGGGAGTGG	217119
QY	481	gaaactaagaatccgatgcgaattaaagtggcgcgcaaatcacgaattagcgaaatat	540
DB	217118	GAAACTAAGAAATCCGATGCAGTTAAAGTTGGCGCGCAAAATPACAGAAATAT	217059
QY	541	gcttcgataaccaacgattcttgactctttaggttaaaactgactctctcgaccttta	600
DB	217058	GCTTCGGATACCAAGCGATCTTGACTCTTTAGGTAAACTGACTTCTTCGACCTCTTA	216999
QY	601	cagactgctctctccaactctgtagcaaaataaacaagcagctgagcttcttaagag	660
DB	216998	CAGGCTGCTCTTCTCCAAATCTGTAGCAAAACAATAACAAAGCAGCTGAGCTTCTTAAAGAG	216939
QY	661	atgcaagataaccagtagtcccgaggaaaaacgctgcaattgctcaattcttaattgat	720
DB	216938	ATGCAAGATAACCCAGTAGTCCCGAGGAAAACGCCCTGCAATTTGCTCAATCTTTAGTTGAT	216879
QY	721	cagacagatgctacgacacacagatagagaagaatggaaaatgcgattggggatgcatat	780
DB	216878	CAGACAGATGCTACAGGCACACAGATAGAGAAAGATGGAATGCGATTAGGCGATGCATAT	216819
QY	781	tttcaggacagaacgctagtggagctgtagaataatgctaaatcttaataacagtaagc	840
DB	216818	TTTTCAGGACAGAACGCTAGTGGAGCTGTAGAAAATGCTAAATCTAAATAACAGTATAAGC	216759
QY	841	aacatagattcagtaaaagcagcaatcgctactgctaacacaaataagctgaagctcag	900
DB	216758	AACATAGATTTCAGTAAAGCAGCAATGCTTACTGCTTAAGACAAATAGCTGAAGCTCAG	216699
QY	901	aaaagtctcccgcactctccaaattcttcaagaacggaacaaatgtaatacaggtcgag	960
DB	216698	AAAAGTTTCCCGACTCTCCAAATCTTCAAGAAGCGGAACAATGGTAATACAGGCTGAG	216639
QY	961	aaagatcttaaaaatatcaaacctcgagatgggtctgtatgttccaaatccaggaactaca	1020
DB	216638	AAAAGATCTTAAAAATATCAAACTGCAAGTGGTTCTGATGTTTCCAATTCAGGAACATCA	216579
QY	1021	gttgagagctccaaacagcaagaaatgattggtatgattctgtgtttccatgctgttta	1080

Db 216578 GTTGGAGGCTCCAAGCAACAGAAAGTAGTATTGGTAGTATTTCGTGTTTCCATGCTGTTA 2165119

QY 1081 gatgagtctgaaaaatgagaccgcttccatttggatgctcgtggttttcgtagatgattcac 1140

Db 216518 GATGATGCTGAAAATGAGACCGCTTCCATTTTGTATGCTGGGTTTCGTGAGATGATTAC 216459

QY 1141 atgttcaatacgaataatcctgattctcaagctgcccacaacagagagctcgcagcacaagct 1200

Db 216458 ATGTTCAATACGGAATAATCTGATTCTCAAGCTGCCACACAGAGAGTCCGACGACAAGCT 216399

QY 1201 agagcagcgaagccgctggagatgacagtgcctgctgcagcgcctggcagatgctcagaaa 1260

Db 216398 AGAGCAGGAAAGCCGCTGGAGATGACAGTGCCTGCTGACGCGCTGGCAGATGCTCAGAA 216339

QY 1261 gctttagaacggccttagtaagctggcgaacacagggcgcatactcaatgctttggaa 1320

Db 216338 GCTTTAGAAGCGGCTCFAGGTAAGCTGGGCAACACAGGGCATACTCAATGCTTTAGGA 216279

QY 1321 cagatcgtctctgctgctgtgtgagcgcaggag-tctctcccgtctgcagcaagtctctatg 1379

Db 216278 CAGATCGCTTCTGCTGCTGTTGTGAGCGCAGGAGTTCTTCCCTCCGCTGCAGCAAGTCTATA 216219

QY 1380 g 1380

Db 216218 G 216218

RESULT 7

AX045135 1456 bp DNA PAT 24-NOV-2000

LOCUS Sequence 5 from Patent WO0066739.

DEFINITION AX045135

ACCESSION AX045135

VERSION AX045135.1 GI:11343734

KEYWORDS Chlamydothila pneumoniae.

SOURCE Chlamydothila pneumoniae.

ORGANISM Bacteria; Chlamydiales; Chlamydiaceae; Chlamydothila.

REFERENCE 1 (bases 1 to 1456)

AUTHORS Murdin,A.D., Omen,R.P., Wang,J. and Dunn,P.

TITLE i(chlamydia) antigens and corresponding dna fragments and uses thereof

JOURNAL Patent: WO 0066739-A 5 09-NOV-2000;

AVENTIS Aventis Pasteur Limited (CA)

FEATURES Location/Qualifiers

source 1..1456

organism="Chlamydothila pneumoniae"

db_xref="taxon:83558"

101..>1452

/note="unnamed protein product"

/codon_start=1

/transl_table=11

/protein_id="CAC17379.1"

/db_xref="GI:11343735"

/translation="MVNPICGPTDERTPPADLSAQGLEASAANKSAEAOIRAGAE

AKPKESKTSVERWSILRSVNALMSLADKLIGTASSNSSSTSRSDVDSTTATPT

PPPTFDYKTAQATYDITFTSLADIQAALVSLQDVAIVNKDTRATDETAIAAE

ETKNADAVKGAITELAKYASDNQAILDSGLKTLFDLLQAALLSVAANNKAELL

KEMQNDVPVKPTPAITQSLVDQTDATATIERKDGNARIDAYFAGNAGSVAENAKS

NSINSDSAKAATATQTAIEAKKFPDPSPIQEAEMOVIQAEKDLKNTKPADSGD

PNPQTTVGGSKQSGSTGSRVSMILLDAENETASILMSGFOMIHMFNTENPDQAA

QQLAAQARAARAKAGDDSAALADAQKALEALGRAGQOQGIILNALGQIASAAVVS

G"

BASE COUNT 452 a 308 c 331 g 365 t

ORIGIN

Query Match 97.0%; Score 1347; DB 6; Length 1456;

Best Local Similarity 99.6%; Pred. No. 3.3e-313;

Matches 1350; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 atggttaatcctattggtccaggctctatagcgaacagaacgcacacctccgcagat 60

	1141	atcttcaataacggaataactgtattctcaagctgccaaacagagcctcgacgacaagct	1200
QY			
Dd			
	1241	ATGTTCAATACGGAAATCCTGATTCTCAAGTCGCCAACAGAGCTCGCAGACAAGCT	1300
Db			
	1201	agaagcgcgaagccgctggagatgacagtgtctgctgcagcgcgtggcagatgctcagaa	1260
QY			
Dd			
	1301	AGAGAGCCGAAGCCGCTGGAGATGACAGTGTGCTGCAGCGCTGGCAGATGCTCAGAA	1360
QY			
Dd			
	1261	gccttagaacgcgctctagtaaacgtggcgaacacaggcatactcaatgctttggaa	1320
QY			
Dd			
	1361	GCTTTAGACAGCGCTCTAGTAAGCTGGCACACAGGGCATACTCAATGCTTTAGGA	1420
QY			
Dd			
	1321	cagatcgcttcgtctgctgttgttgagcgcaggagt	1355
QY			
Dd			
	1421	CAGATCGCTTCTGCTGCTGTGTGTGAGCGCAGGAGT	1455
QY			
Dd			
	RESULT 8		
	AXO45133		
LOCUS	AXO45133	1852 bp DNA	PAT 24-NOV-2000
DEFINITION	sequence 3 from Patent WO0066739.		
ACCESSION	AXO45133		
VERSION	AXO45133.1	GI:11343732	
KEYWORDS			
SOURCE	Chlamydomophila pneumoniae.		
ORGANISM	Chlamydomophila pneumoniae		
REFERENCE	Eacteria; Chlamydiadiales; Chlamydiaceae; Chlamydomophila.		
AUTHORS	Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.		
TITLE	-i(chlamydia) antigens and corresponding dna fragments and uses thereof		
JOURNAL	Patent: WO 0066739-A 3 09-NOV-2000;		
FEATURES	Aventis Pasteur Limited (CA)		
Source	Location/Qualifiers		
	1..1852	/organism="Chlamydomophila pneumoniae"	
	/db_xref="taxon:83558"		
CDS	1..1752	/note="unnamed protein product"	
	/codon_start=1		
	/transl_table=11		
	/protein_id="CACI17378.1"		
	/db_xref="GI:11343733"		
	/translation="MSLADKLGIASSNSSSTSRSDVDSTTATPTPPPTFDYKT QAQTADTFITSLADIQAALVLQDAVTNKDTAATDEETAIAAEWTKNADAVK GAQTELAKYASDNGAILDLSGLKTLFPLLQNALLQSVAANNKAELKEMODPVP GKTPAIAGSLVDOTATATIEKDGNARIDAYFAGNQSAVENAKSNNSISNDISA AAIATQTOIAEVRQFPDSPILQAEQVIQAEKDLINKIPADGSVPNPQTVVGGS KOQSGSIIRVSMULDNAENETASILMSGFOMIHMETPDSPQAQOEILAAQARA AKAGDDAAAALADAQRALKAAGKAGQQOQILNALQIIASAIVSVAGVPPAAASSI GSSVOLYKTSKSGDYKTOIGSAGDAYKSIDNAYGRANDATRDVINNVSTPALTRI SVPRATREARGPKTDQALRVISGNSRTLGDVYSQVSALQSVMDOITOSNPOANNEEI RQKLTSVATKPPEQGPYPVOLNSDSTQKFIAKLESIFABEGSTAEEIKALSFPETNSLF IQOVLYNIGSLSYGLQ"		
BASE COUNT	578 a 409 c 407 g 458 t		
ORIGIN			
Query Match	83.2%; Score 1155.4; DB 6; Length 1852;		
Best Local Similarity	99.4%; Pred No. 3.5e-267;		
Matches 1170; Conservative	0; Mismatches 6; Indels 1; Gaps 1;		
QY	205	atgagtctggcagataagctgggatgtgcttctagtaaacagctcgctcttctactagcaga	264
Db			
	1	ATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTTAGTAACAGCTCGCTCTTACTAGCAGA	60
QY	265	tctgcagactggactcaacgacagcagcaccttacgctctctcccaccagctctgat	324
Db			
	61	CTTGACAGCTGGACTCAACGACGACGACCGACCCTACGCGCTCTCCACCCAGTTTGAT	120
QY	325	gattataagactcaacgcgaacacagcttacgatactattcttacctccaacatacctagct	384

Db 121 GATTATAGACTAAGCGCAACAGCTTACGATACCTATCTTTTACCTCAACATCACTAGCT 180
Qy 385 gacatacaggctcttggtagcctccagatctctcaataataaagatacagc 444
Db 181 GACATACAGGCTGCTTTGGTAGCCTCCAGATGCTGTCTACTAATATATAAGGATACAGCG 240
Qy 445 gctactgatgagaaaccgcaatcgctgcgagtgaggaaactaagaatgcccagtgcaatt 504
Db 241 GCTACTGATGAGAAACCGCAATCGCTCGGAGTGGGAACATAGAAATGCGGATGCACTT 300
Qy 505 aaagtggcgcgcaaatatacagaattagcgaataatgcttctggataaccagcattctt 564
Db 301 AAAGTTGGCGCGCAATACAGAAATAGCGAAATATGCTTCGGATAACCAAGCGATTCTT 360
Qy 565 gactcttagtaaacgactctctcgactcttcaagactctcttcccaatctgta 624
Db 361 GACTCTTTAGTAACAGTCTCTTGGACCTCTTACAGGCTGCTCTTCTCCAATCTGTA 420
Qy 625 gcaacaataaacaagcagctgagctctttaaagagatgaagataacccagtagtccca 684
Db 421 CCAACATAACAAGCAGCTGAGCTCTTAAAGAGATGCAAGATAACCCAGTAGTCCCA 480
Qy 685 gggaaacgcgcgaattgctcaatctttagttgatcagacagatgctacagcagacag 744
Db 481 GGGAAACGCTGCAATTCCTCAATCTTTAGTTGATCAGACAGATGCTACAGCGACACAG 540
Qy 745 atagagaagatgaatcgatgggagtcataatttttcagacagacagcagtaggga 804
Db 541 ATAGAGAAGATGGAATGCGGATGAGGATGCAATTTTGCAGGACAGAACGCTAGTGGG 600
Qy 805 gctgtagaaatgctaaatcctaataacagataaagcaacatagattcagctaaagcagca 864
Db 601 GCTGTAGAAATGCTAATCTAATAACAGTATAAGCAACATAGATTACAGTAAAGCAGCA 660
Qy 865 atcgcactgtaagacacaatagctgaagctcagaataagttcccgactctcccaatt 924
Db 661 ATCGTACTGTAGACACAAATAGCTGAAGCTCAGAAATAAGTTCCCGACTCTCCAAAT 720
Qy 925 ctcaagaagcgaacaatgtaatacagctgagaagaatcttaaaatcaaacct 984
Db 721 CTTCAAGAAGCGGAACAATGGTATACAGGCTGAGAAGATCTTAAATAATCAAACT 780
Qy 985 gcagatggtctgattgccaaatccaggaactacagttgaggtcccaagcaacaagga 1044
Db 781 CGAGATGTTCTGATGTTCCAAATCCAGGAACCTACAGTTGGAGGCTCAACGAACAAGGA 840
Qy 1045 agtagtattgtagtattcgtgtttccatgctgttagatgctgtaaaatgagaccgt 1104
Db 841 AGTAGTATTGTTAGTATTCTGTTTCCATGCTGTGTAGATGATGCTGTAATAATGAGCCGCT 900
Qy 1105 tccatttgatgtcgtgttcgtagatgattcacatgtttcaatcaggaataatcctgat 1164
Db 901 TCAATTTGATGTCGTGGGTTTCTGATGATGATTCATGTTCAATACCGGAATTCCTGAT 960
Qy 1165 ttcaagctgcccaacagagctgcagcagcaagaagctagagcagcaagccgctggagat 1224
Db 961 TCTCAAGCTGCCAACAGAGCTCCAGCAACAAGCTAGAGCAGCGAAAGCGCTGGAGAT 1020
Qy 1225 gacagtgtcgtgcagcgtgagcagatgctcagaagaatttagaagcgtctcaggtaaa 1284
Db 1021 GACAGTGTGTGTCAGCGCTGCGAGATGCTCAGAAAGCTTTAGAAAGCGGCTCTAGGTAAA 1080
Qy 1285 gctggcgaacacagggcatactcaatgctttggacagatcgctctgctgctgttg 1344
Db 1081 GCTGGCAACACAGGGCATACTCAATGCTTTAGGACAGATCGCTTCTGCTGCTGTGTG 1140
Qy 1345 agcgaggag-tcctcccgctgcagcaagttctatgg 1380
Db 1141 AGCGCAGGAGTTCTCTCCCGCTGCAGCAAGTCTATAG 1177

RESULT 9
AR144067

LOCUS AR144067 150 bp DNA PAT 08-AUG-2001
DEFINITION Sequence 8 from patent US 6210876.
ACCESSION AR144067
VERSION AR144067.1 GI:15105934
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 150)
AUTHORS Cerney, M.B.
TITLE Nucleic acid primers and probes for detecting Chlamydia pneumoniae
JOURNAL Patent: US 6210876-A 8 03-APR-2001;
FEATURES Location/Qualifiers
1..150
source
BASE COUNT 43 a 34 c 40 g 33 t
ORIGIN
Query Match 10.6%; Score 146.8; DB 6; Length 150;
Best Local Similarity 98.7%; Pred. No. 1.1e-24;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 366 tacctcaacatcactagctgacatacaggtgctttgtgtgagctccaggtgtgtcac 425
Db 1 TACCTCAACATCACTAGCTGACATACAGGCTGCTTTGGTAGCCTCCAGGATGCTGTGAC 60
Qy 426 taatataaagatcacagcgctactgatgagaaacccgcaatcgtcggagtggaaac 485
Db 61 TAAATATAAGGATACAGCGGCTACTGTAGTGAAGAACCCCAATCGCTGGGTGTGGGAAC 120
Qy 486 taagaatgccgatgcaattaaagtggcgc 515
Db 121 TAAGATGCCGATGCAGTTAAGTTGGCGC 150
RESULT 10
AE001333/c
LOCUS AE001333 10954 bp DNA BCT 30-OCT-2000
DEFINITION Chlamydia trachomatis section 60 of 87 of the complete genome.
ACCESSION AE001333 GI:101273
VERSION AE001333.1 AE001333.1
KEYWORDS
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis
REFERENCE 1 (bases 1 to 10954)
AUTHORS Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V. and Davis, R.W.
TITLE Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
JOURNAL Science 282 (5389), 754-759 (1998)
MEDLINE 99000809
PUBMED 9784136
REFERENCE 2 (bases 1 to 10954)
AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R.W.,
Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
JOURNAL Nat. genet. 21 (4), 385-389 (1999)
MEDLINE 99206606
PUBMED 10192388
REFERENCE 3 (bases 1 to 10954)
AUTHORS Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V. and Davis, R.W.
TITLE Direct Submission
JOURNAL Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
FEATURES Location/Qualifiers
1..10954
/organism="Chlamydia trachomatis"
/strain="D/UW-3/CX"

```

/gene
complement(146. .2089)
/genes="CT622"
CDS
complement(146. .2089)
/genes="CT622"
/codon_start=1
/transl_table=11
/product="CHLPN 76kDa Homolog"
/db_xref="GI:3329069"
/transl_table=11
/translation="MESGPESVSSNQSSMNPPIINGQIASNETKSTKESEASPSASS
SVSFWFLSNAKRALISLRDAILKNSSPTDSLSQLEASTSTSTVTTRVAARDYNEAKS
NFTAKSGSARALLAEYETKMDALQDMERLAKQAEVTRIKALQEKQEVIDK
LNQVLKLEKQNTLKTETTLTTSADQIPAINSOLETKNSAQDQIILKDEQONISYEA
LTVAGVEIKRASSAGIKGOALQSIVDAQSOAAVLQAOONSPDNIAATKLIDAA
ETVKNELQKHTGLTDSPLVKAEQISQAOKDIOETKPSGSDIPVPGSGSAASAGS
AVCALKNSSNGRISILLDDVDNEMRAIAMQGRFSRIEQFNVNPNATAKLOAMEAQL
TAMSDQLVAGDGLPRAETQAIKALQAQLKQPSDGLATANGQVAFAAKLVGGSGAGT
AGTVQMNKOLYKTAFTSSSSSYAALSQGYSAIKTNSLYSESRSQVQSAISQTNAN
PALRSVSRSGIESQSRSDASQRAAETIVRDSOTLGDVYSRLQVLDLSMSTIVSNPQ
VNOEEMQKLTASIKAPQFGYPQVQNSADSLQFAAQLEREFVQDGRSLAESRENAF
RKQPAFTQOVLVNIASLFGYLS"
complement(2197. .3495)
/genes="CT623"
complement(2197. .3495)
/genes="CT623"
/codon_start=1
/transl_table=11
/product="CHLPN 76kDa Homolog"
/db_xref="GI:3329072"
/transl_table=11
/translation="MKKYIKFVGVGALLACAGTSLNLAFAQAASSMDSQSWVEDLDSYL
SSKGFVETRRKQVLRAGDVRAWRIYAKEDLETTOTPAKPMPLPTNRYRSEFNLYVDY
TAANSWTKMNNVTIAGGESSAAGLDINRAFLGYRKYKNPETOAEVFAIGRSLGD
IPDSQVQNSNPDGIIHYAARRISEKLPPTMIHVGPGFVNVMAEKVAVVVEAILNKL
PGNFVKTSVVDNHTLAKTNDPADASAAQAPAKPNKYDLYLVQWLVKGTSTAMPWNG
QTKNLYTGAYLNPDAEPENKQSTPTTKITNGKENHAWFICGSLGVRRAQDWS
ATVRYEVEALAIPEIDVAGIGRGNQMYFAQAIKQGLDPKESNGFTNTYKGVSYQFV
MGTVDSFRAYAAYSKPNANDNLGSDFTYRKYDLGLISS"
complement(3738. .5348)
/genes="mvln"
/notes="CT624"
complement(3738. .5348)
/genes="mvln"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/db_xref="GI:3329071"
/transl_table=11
/translation="MRKDEGSLVRSFLNLLSGTFFSRFLGMLREIVMATYFGADPLV
ASFLAFRTIFFRLKLGPIGLGAFIPHEFLRAQNISRAFTFFRSFRFCYSAIL
FTLIIELGLVCWCVSGTSLFDLTLLTIIILPSGIFLMYTVNSTLHCEKKEFSVGL
APSVNVNWTGTVFLARNYDPRNRIFGLAVLVIGFLEWAVTLPGVMKFLGOSKEVP
QERDSRALIAPLSGLISMGIFQLNLLCDMLARYINEVGPPLYLAYSRIQQLPVHL
FGLGVFTLLPALSRQVQDEHQGDLLRFLSKLVAVNVMTWGLLLPALPGVRVL
YENGVPFKTAHVEIVRGISGIIIPMALAPLVSAFVARRNYKVMVLVGIIAAVVN
MVLNVJGLCVKQVAVLTYATVSLVSGQLAMLYCAGKSPTYKGLMWRTFKESGKTV
ITVLAATVIGNVITHYTVVYFIEPLTVPTKPLVSLFDQCGVFAESALFLSVLFG
LAKLLKEDLMLNLISFYWKGHQSILRN"
5521. .6387
/genes="nfo"
5521. .6387
/genes="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AAC68229.1"
/db_xref="GI:3329072"
/transl_table=11
/translation="MFLPPQFALLGAHTSAAGLHNLAYEGRIDGATTVOLFTANQ
RQKRRRTLQEWDPQAFRIALNLSLVSIMSHAGYLNPNPAPNEILKTKVCMHOETA
DCISLGISFVNHPHGAALSDSKESCLDRITASFQMAPLFENNPPLVLLLETTAGGGS
LIGSSFEELAYLQGIKAHPIGVCLDTCHIFPAAGYDISSVAGVEQVLKHFDAVIGLS
FLRAIHLNDSVFALGKGNKRRHAPIGEGCIGSDSPCFMLQDQTRMLPKYLETGGPDL
WTKEIRYLQKVC"
complement(6484. .7113)
/genes="rs4"
complement(6484. .7113)
/genes="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AAC68230.1"
/db_xref="GI:3329073"
/transl_table=11
/translation="MARYCGPKNRIARRFGANIFGRGNPLLRPNPPGQGMQMKKK
SDYGLQEEOKLCAKGMILEKQVLKAYEVNKNQVNAOMLEKTECRLDNIVYRL
GFAKTIFAAOLVSHGHVNLVNGKVDRRSPFVRPMOISLKEKSKRLAIVTESLENKD
OSSLPAYLSLDKAAFKGELVVAPELDQIASOLPLPVNVSVICEFLSHRT"
7497. .8480
/genes="ycea"
7497. .8480
/genes="ycea"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68231.1"
/db_xref="GI:3329074"
/transl_table=11
/translation="MEKNYVALVYFVPVSNPHEETALHKOLFCTMDVSCRIYISEE
GNGQESGYOPDAERYMAWLKQRFDAISIKFIHIEENTFPRTVYKRELVALGCS
VDTTKQKHISPEWHEKRLQENCLVLDVRNNYEMKIGHFENAVLPDITFPFPDIA
DLRAKHDPARTPVMYCTGIGRCLELYSALLKEGFKEVQLDGVGIAYGLKMTGKW
RKLGFVDDRRAMPIDEADPNVSPAIARCSLCNTSDTYINCANTDCNLFTECCESCIA
THKGCSSECSQAPRIADFSAERGNKPFRRKHLCTPTIEQSCCLKEQENQPA"
complement(8552. .9427)
/genes="ispa"
/notes="CT628"
complement(8552. .9427)
/genes="ispa"
/codon_start=1
/transl_table=11
/product="Geranylgeranyl pyrophosphate synthase"
/protein_id="AAC68232.1"
/db_xref="GI:3329075"
/transl_table=11
/translation="MDYFVSIVRSVKKLRDSGLDQFNTGGLRDPIEYALLGGKRV
RPLVLGFAEGVHKHREDVLDTAIEYIHTSTLIADDLPCMDDDDDMRGRPSVHKAFD
EASALLASALIPAAAYARIKNAKALKAVVSCQEDVEEAYEDILELIEFRGVGILG
GOYEDIFFHDFSKENVLGIIKKKTGALFEIACVSGWLFGGGERESSTLVAFEAHFGL
LFOIRDDLADLSQDDQOEKHMNYALLFGEVAAKELLDSHESFCIKNLHLLQERGLECL
EPLMLCKNVFCGWK"
complement(9483. .10100)
/genes="glmU"
complement(9483. .10100)
/genes="glmU"
/codon_start=1
/transl_table=11
/product="UDP-GlcNAc Pyrophosphorylase"
/protein_id="AAC68233.1"
/db_xref="GI:3329076"
/transl_table=11
/translation="MVLSSLSFSEPEELYPEIVSOAEFVMSILTLLEKLAHSTFSGI
HGHLBEGVYLKNETIEIQGAYVESGAYICGPCIIGPYQVRHGAIVRGVITSSHC
VIGHSEIKNSYLGHHAKAAHFAVYGVDSVSRNGLAGVRCANFRDLGKTIFFHHS
ERCNTRKRLKGAFLGRGVSGVNCNTVLPNPGCYVASATKILPNQTIY"
complement(10153. .10836)
/genes="cpxR"
/notes="CT630"
complement(10153. .10836)
/genes="cpxR"
/codon_start=1
/transl_table=11
/product="HTH Transcriptional Regulatory Protein and Receiver Domain"
/protein_id="AAC68234.1"
/db_xref="GI:3329077"

```

/translation="MAGPKHLLVSENWDLFFOTKELLNPEYRCITGOOYKOELSDAD
LVVCEYSLPREVSKPSKLVLLDFDEETSDVLLDRGFVLIQIPITRILKS
AISLFLSOLSHSVPSIRPGPNVHLKLTVPESVHLTPSESGILKRLINKGO
LCLKKHLEIEIKHAKAIVARNVDVHLASLKKLGLAYGSRIVTLRGVYLFSDGDKK
FSOODTKLS"
BASE COUNT 3343 a 2463 c 1958 g 3190 t
ORIGIN

Query Match 4.1%; Score 56.6; DB 1; Length 10954;
Best Local Similarity 46.0%; Pred. No. 0.0039;
Matches 406; Conservative 0; Mismatches 444; Indels 33; Gaps 5;
Qy 506 aagttgagcgcaaatcacgaattagcgaataatgcttcgataacacgaattcttg 565
Db 1623 AGGTATTGTAAGCTCANTCAGTTAGTTAACTTGAACACGAATCAGACTTTAAAGG 1564
Qy 566 actcttaggttaactgacttccttcgacctcttcacagactgctcttccaattctgtag 625
Db 1563 AAACCTTTAAACAACACAGACTCTGCAGATCAGATTCAGCGATTAAATAGTCAGTTAGAGA 1504
Qy 626 caacaataacaaagcagctgagcttcttaagagatgaagataaacccagtagtcccag 685
Db 1503 TCACAAATAATCTGCAGATCAATATCAAAAGATCTGGAAGGACAAACACATAAGTTATG 1444
Qy 686 ggaacacgctgcgaatgctcaatcttagttgatcagacagatgctacagcgacacaga 745
Db 1443 AAGCTGTTCTCATAACGACGAGAGGTTATCAAAAGCTTCTTCGAAAGCGGGAATTAAGT 1384
Qy 746 tagagaaagatggaaatgctgagtgatgcatattttgcaggacagacgctagtgag 805
Db 1383 TAGGACAAGCTTTGCAGCTTATTGTTGATGCTGGGATCAAGCCAGGCTCAGTTCTTC 1324
Qy 806 ctgtagaaatgctaaatcctaataacagataaagacacatagattcagctaaacagcaa 865
Db 1323 AAGCACAGCAAAATAATACCCAGATATATCGCAGCACCAAGAAATAATTAATGATGCTG 1264
Qy 866 tcgctactgtaagacacaaatagctgaagctcagaaaaagttcccgactctccaatc 925
Db 1263 CTGAACGACGTAACGAGTTAAACAAAGACATACAGGCTAACGGACTCGCCTTTAG 1204
Qy 926 ttcaagaagcggaacaaatgtaatacaggtgagaaagattcttaaaatatcaaacctg 985
Db 1203 TCAAAAGGCTGAGGAGCAGATTAAGTCAAGCACAAGAAAGATTAATCAAGAGATCAAACT- 1143
Qy 986 cagatggtctgattgtccaaatccagaaacta--cagttgaggtctccagcaacaaagg 1043
Db 1144 --AGTGGTTCGGATATTCCTATCGTTGTCGAGTGGTTCAGCTGCTTCCGACGGAAGTG 1087
Qy 1044 aagtagtattgtagtattc-----gtttccatgctgtagtagtg 1087
Db 1086 CGGTAGGAGCGTTGAATCCTTAACAAATTCAGGAAGAAATTCCTTGTGCTTGTATGATG 1027
Qy 1088 ctgaaaaatgagacgctccattttagtctgtggtttcgcagatgattcacatgttca 1147
Db 1026 TAGCAATGAATGCGCAGCGATTGCAATGCAAGGTTTTCGATCTATGATCAACAAATTA 967
Qy 1148 atacggaataatccgattctcaagctcccaacagagagctcgcagcaacagctagagcag 1207
Db 966 ATGTAAACATCCTGCAACAGCTTAAGAGCTTACAGCTATGGAGCTCAGCTGA---CTG 910
Qy 1208 cgaagcgctggagatgacagtgctgctgagcgctgagcgatgctcagaaagctttag 1267
Db 909 CGATGTCAGATCAACTGTTGTCGGATGCGGATGCGAGCTCCAGCGCAATACAGCAATCA 850
Qy 1268 aagcgctctaggtaagctgggcaaac-----agggcatactcaatgcttgg 1318
Db 849 AAGATGCTCTTTCGCAAGCTTGAACAAACCAACAGATAGGTTTACGTACAGCTATGG 790
Qy 1319 gacagatgctctgctgcttattgtgagcgagagctctccc 1361
Db 789 GACAAGTGGCTTTGACAGCTGCCAAGGTTGGAGGAGGCTCCGC 747

RESULT 11
166494/c
LOCUS
DEFINITION Sequence 14 from patent US 5670367.
ACCESSION 166494
VERSION 166494.1 GI:2724471
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 7218)
AUTHORS Dörner, F., Scheiflinger, F. and Falkner, F. Gunter.
TITLE Recombinant fowlpox virus
JOURNAL Patent: US 5670367-A 14 23-SEP-1997;
FEATURES Location/Qualifiers
source 1..7218
BASE COUNT 1944 a 1491 c 1486 g 1929 t 368 others
ORIGIN

Query Match 3.9%; Score 54.4; DB 6; Length 7218;
Best Local Similarity 9.3%; Pred. No. 0.014;
Matches 43; Conservative 218; Mismatches 199; Indels 0; Gaps 0;
Qy 631 aatacaaaagcagctgagcttcttaagagatgcaagataaacccagtagtcccagggaaa 690
Db 1481 AATTACCTTATTCGCAAGTAGTTAAAGAGATAGAGAATTTGGTACRRRRRRRRRRR 1422
Qy 691 acgctgcaattgctcaattcttagttgatcagacagatgctacagcgacacagatagag 750
Db 1421 RRR 1362
Qy 751 aaagatgaaatcgattgggagtcataatttcgagcagacaaacgctagtgagctgta 810
Db 1361 RRR 1302
Qy 811 gaaatgctaaatcaataacagtaataagcaacatagattcagctaaagcagcaatcgct 870
Db 1301 RRR 1242
Qy 871 actgtaagacacaaatagctgaagctcagaaaaagttcccgactcccaattctcaa 930
Db 1241 RRR 1182
Qy 931 gaagcggaacaaatggtaatacaggtgagaaagatcttaaaatatcaaacctgcagat 990
Db 1181 RRR 1122
Qy 991 ggttctgattctccaaatcccgaaactacagttgagctcgaagcaacaggaagtagt 1050
Db 1121 RRR 1062
Qy 1051 attgtagtattcgtttccatgctgtagtagtagtgcgtg 1090
Db 1061 AAGTCCCTCGACCTGCGCAAGCTCGGAATTAATTCG 1022

RESULT 12
AX155814
LOCUS
DEFINITION Sequence 57 from Patent WO0140474.
ACCESSION AX155814
VERSION AX155814.1 GI:14536973
KEYWORDS
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis.
REFERENCE 1 (bases 1 to 1537)
AUTHORS Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
TITLE Probst, P., Bhatia, A., Skeiky, Y. A., Fling, S. P. and Scholler, J. Compounds and methods for treatment and diagnosis of chlamydial infection

Qy 922 attcttaagaacggaacaaatggtaatcacagcctgagaagatctttaaataatcaca

Search completed: February 7, 2002, 19:33:50
Job time: 13544 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 12230.8 Seconds
(without alignments)
3432.748 Million cell updates/sec

Title: us-09-391-606-1
Perfect score: 2545
Sequence: 1 ttgcggtgtgttaacggtg.....tgccactccactgtctctt 2545

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_om.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vl.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_om.*
20: em_or.*
21: em_ov.*
22: em_pat.*
23: em_ph.*
24: em_pl.*
25: em_ro.*
26: em_sts.*
27: em_sy.*
28: em_un.*
29: em_vl.*
30: em_htgo_hum.*
31: em_htgo_inv.*
32: em_htgo_rod.*
33: em_htg_hum.*
34: em_htg_inv.*
35: em_htg_rod.*
36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No	Score	Match	Length	DB	ID	Description
1	2225.2	87.4	2238	6	AX045137	AX045137 Sequence
2	2122.2	83.4	10421	1	AE001654	AE001654 Chlamydia
3	2122.2	83.4	12548	1	AE002165	AE002165 Chlamydia
4	2122.2	83.4	300550	1	AP002547	AP002547 Chlamydia
5	2109.2	82.9	2148	1	CHT76KDA	L23921 Chlamydia p
6	1459.4	57.3	2156	6	AX045131	AX045131 Sequence
7	1447	56.9	1456	6	AX045135	AX045135 Sequence
8	1155.4	45.4	1852	6	AX045133	AX045133 Sequence
9	264	10.4	14698	1	AE002357	AE002357 Chlamydia
10	257	10.1	10954	1	AE001333	AE001333 Chlamydia
11	217.6	8.6	3455	1	CTU50732	U50732 Chlamydia t
12	200.2	7.9	6849	12	AY029367	AY029367 Synthetic
13	171	6.7	5108	6	AX138932	AX138932 Sequence
14	171	6.7	5108	6	AX138933	AX138933 Sequence
15	171	6.7	5111	6	AX138936	AX138936 Sequence
16	171	6.7	5185	6	AX138928	AX138928 Sequence
17	171	6.7	5188	6	AX138940	AX138940 Sequence
18	171	6.7	5254	6	AX138941	AX138941 Sequence
19	149.2	5.9	4928	6	AR141727	AR141727 Sequence
20	149.2	5.9	4928	6	I49890	I49890 Sequence 1
21	146.8	5.8	150	6	AR144067	AR144067 Sequence
22	145	5.7	7475	6	AX001323	AX001323 Sequence
23	145	5.7	7475	6	E36259	E36259 Human semap
24	145	5.7	8192	6	AX001324	AX001324 Sequence
25	145	5.7	8192	6	E36260	E36260 Human semap
26	142	5.6	3125	6	AR034783	AR034783 Sequence
27	142	5.6	7108	6	AX001326	AX001326 Sequence
28	142	5.6	7108	6	E36262	E36262 Human semap
29	139.8	5.5	4276	6	AX154496	AX154496 Sequence
30	139.8	5.5	4282	6	AX164583	AX164583 Sequence
31	139.8	5.5	4328	6	A92666	A92666 Sequence 2
32	139.8	5.5	4328	6	AR137192	AR137192 Sequence
33	139.8	5.5	4818	6	A92668	A92668 Sequence 3
34	139.8	5.5	5107	6	A92667	A92667 Sequence 4
35	139.8	5.5	9600	6	A92665	A92665 Sequence 1
36	139.2	5.5	5459	6	AX154499	AX154499 Sequence
37	139.2	5.5	5882	6	AX154501	AX154501 Sequence
38	139.2	5.5	7731	6	A94048	A94048 Sequence 29
39	139.2	5.5	7731	6	A94061	A94061 Sequence 42
40	139.2	5.5	7731	6	AX011133	AX011133 Sequence
41	139.2	5.5	7731	6	AX011146	AX011146 Sequence
42	139.2	5.5	8068	6	A94046	A94046 Sequence 27
43	139.2	5.5	8068	6	A94054	A94054 Sequence 35
44	139.2	5.5	8068	6	AX011131	AX011131 Sequence
45	139.2	5.5	8068	6	AX011139	AX011139 Sequence

ALIGNMENTS

RESULT 1

AX045137

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

source

PAT

24-NOV-2000

DNA

Sequence 7 from Patent WO0066739.

AX045137

AX045137.1 GI:11343736

Chlamydia pneumoniae.

Chlamydia pneumoniae

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

1 (bases 1 to 2238)

Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.

-(Chlamydia) antigens and corresponding dna fragments and uses thereof

Patent: WO 0066739-A 7 09-NOV-2000;

Aventis Pasteur Limited (CA)

Location/Qualifiers

1..2238

/organism="Chlamydia pneumoniae"

/db_xref="taxon:83558"

```

CDS      766..2238
        /note="unnamed protein product"
        /codon_start=1
        /transl_table=11
        /protein_id="CAC17330.1"
        /db_xref="GI:11343737"
        /translation="MVNPVIGPGPIDETPPADLSAQGLEASAAKSAEQAQRIAGAE
        AKPKESKTSVVERWSILRSVAVNALMSLADKLGIASSNSSSTSRSDVDSTTATATPP
        PPPTFDYKTKQAOPAYDTIFSTSLADIQALVSIODAVTNIKDTATDETAIAEAW
        ETKNADVKVQAQITELAKYASDNQAILDSGLKTSFDLLQALLQSVANNKRAEUL
        KEMDNVPVPGKTPAIASLVQDQDATQIEKDGNARIDAYFAGNAGAVENAKNS
        NSINIDSAKAAIAITAKTQIAEAKKFPDPIQEAQOMVIOAEKDLNKPADGSDV
        PNPQTTVGRKQKQSDSSIRVSMLLDNEATASILMSGFROMIHMFENTENPDSQAA
        QOELAAQAKAAKAGDSSAAAALADAKALEAALGKAGQOQILNALGQIASAAVWSA
        GVLPQQVLIRAYQAVVEQKLISEEDLNSAVDHHHHH"
BASE COUNT 689 a 459 c 497 g 593 t
ORIGIN

Query Match      87.4%; Score 2225.2; DB 6; Length 2238;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2230; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 175 atgacaaaaacattatgcttgggtgtagaaggattctcaatcggtttgcctaaacag 234
DB 1 ATGACAAAAACATATGCTTGGGTGTAGAAGGATTCTCAATCGTTTGCCTAAACAG 60

QY 235 ttttttgtaaatgtagtgtgtcagatggaacacattcggttcttcagaacctccact 294
DB 61 TTTTGTGTAATGTAGTGTGCTGACATGGAACACATTCGTTCCTCAGAACCTCCACT 120

QY 295 acagaaaaagctgtacaaaagctatgaatacaaaactgtgtttggcagtgctcgtc 354
DB 121 ACAGAAAAAGCTGTACAAAAGCTATGAATACAAATACTGTGTTGGCAGTGGCTCGTC 180

QY 355 gaaagacatagtcaggttcttggatcaatggacaaaaagcctcttatctttatgga 414
DB 181 GAAAGACATAGTCAGGTTCTTGGATCAATGGACAGAAAAAGCCTCTATATCTTTATGGA 240

QY 415 gcttcttaatagaacctttagcaagctcagaagactacgttaataatggaaaaagaaac 474
DB 241 GCTTCTTAAATGAACCTTTTAGCAAGCTACGAAAGACTAGCTTAAATGGAAGAAAAAC 300

QY 475 ctagcttggtttatggagaaactttaggggactcagaaaagctggagactggtctgcc 534
DB 301 CTAGCTTGGTTATTTGGAGAACTTTAGGGGACTCAGAAAAGCTGGAGACTGGTCTGCC 360

QY 535 acagtcagttatgtagtgcgaagccttgcagttccagaaaatagatgttttcagggtat 594
DB 361 ACAGTACGTTATGATGTGCAAGCCTTGTGCGGTTCCAGAAATAGATGTTTCAGGGATT 420

QY 595 ggcctgtgtaatttataaagtttgggtgcgcaagaactgtgcttaactatgatcct 654
DB 421 GGCCTGTGTAATTTTAAAGTTTGGTTCGCCCAAGCAATTCGTCTTAACATATGATCCT 480

QY 655 aaagaggcctaagttttacaaaattataaaggattttccgctctatatatgatggcacc 714
DB 481 AAAGAGGCTAATGGTTTACAAATATTAAGGATTTCCGCTCTATATGATGGCAATC 540

QY 715 acagattctctatcattcagagcttatgggcttactccaaaccagcaaaacgaataaactc 774
DB 541 ACAGATTCTCTATCATTACAGACTTATGGGCTTACTCCAAACACAGCAACGATAAATC 600

QY 775 ggcaggtatttacttccgaagtttgatctaggtataaattcagcgttttaagtcaaa 834
DB 601 GGCAGTGTATTTTACTTCCGAAGTTTGAICTAGGTATAAATTTACGCGCTTTTAAGTCAA 660

QY 835 ttttaataaaactcttaaaacagctcgcatattattagtgagagctttttttat 894
DB 661 TTTTAATAAAATCTTTAAACACAGCTCGCATTAATTAATAGTGAGAGCTTTTTTTTTAT 720

QY 895 tttttataaaaaactaaaagattttatttttttgagtttttttttatggttaactctatt 954

```

```

DB 721 TTTTATAATAAACTAAAGATTTTATATTTTGGTGTATGTTTAACTCTATT 780
QY 955 ggtccaggtcctatagacgaaacagacacactcccgagatctttctgctcaagga 1014
DB 781 GGTCCAGGTCTATAGACGAAACAGACACCTCCCGCAGATCTTCTTCGCTCAAGGA 840
QY 1015 ttgagggcagtgagcaaaaataagagtcgaggaagctcaaaagatagcaggtgcggaagct 1074
DB 841 TTGGAGGGGAGTGAGCAAAATAAGAGTGGGAAGCTCAAGAATAAGCAGGTGCGGAAGCT 900
QY 1075 aagcctaaagaatctaagaccgattctgtagagcagtgagcattctgtctgcagtg 1134
DB 901 AAGCCTAAAGAATCTAAGACCGGATTCTGTAGAGCGATGGAGCATCTTCGCTTCGCACTG 960
QY 1135 aatgctctcatgagctgcagataaagctgggtattgcttcttagtaaacagctgctctct 1194
DB 961 AATGCTCTCATGAGTCTGGCAGATAAAGCTGGGTATTTCTTAGTAACAGCTCGCTTCT 1020
QY 1195 actagcagatctgcagacgttgactcaacgacagcgcacacctacgctctctccccc 1254
DB 1021 ACTAGCAGATCTGCAGAGCTGGACTCAACGACAGCGACCGACCTACGCTTCTCCACCC 1080
QY 1255 acgtctgatattataagactcaagcgcgaacagcttacgatactattttacotcaaca 1314
DB 1081 ACGTTTGATGATTATAAGACTCAAGCGCAACAGCTTACGATACCTATCTTTACCTCAACA 1140
QY 1315 tcaactagtcacatacagagctgcttggtagcctcagagtgctgtcactaataaag 1374
DB 1141 TCACGTAGCTGACATACAGGCTGCTTGGTGGCTCCAGGATGCTGCACATAATAAG 1200
QY 1375 gatacagcggctactgatgagaaacccaactcgtcgaggtgggaactaaagaatgcc 1434
DB 1201 GATCAGCGGCTACTGATCAGAAACCCCAATCGCTGCGGAGTGGGAACCTAAGNAAGCC 1260
QY 1435 gatcaattaaagttggcgcgaataattacagaattagcgaataatgcttcggataaccaa 1494
DB 1261 GATGCAGTAAAGTTGGCGCAAAATTACAGAAATTAGCGAAATATGCTTCGATAAACCA 1320
QY 1495 gcgattcttgactcttttagttaactgaactccttcgacctcttacagactgctctctc 1554
DB 1321 GCGATTCCTGACTCTTTAGGTAACCTGACTTCTTCGACCTCTTACAGGCTGCTCTTCTC 1380
QY 1555 caactctgtagcaacaataacaaagcagctgagctctttaaagagatcaagaataaccca 1614
DB 1381 CAATCTGTAGCAACAATAACAAACCACTGAGCTTCTTAAAGAGATGCAAGATAACCCA 1440
QY 1615 gtagtcccgaggaaaacgcctgcgaattgctcaatttttagttgatcagacagatgctaca 1674
DB 1441 GTAGTCCCGAGGAAAAACGCTGCAATTCGCTCAATCTTTAGTTGATCAGACAGATGCTACA 1500
QY 1675 gcgacacagatagagaaagatggaaatgcgattggggatgcatatttttgcaggacagaaac 1734
DB 1501 GCGACACAGATAGAGAAAGATGGAAATCGGATTAGGGATGATATTTTTCAGAGACAGAAC 1560
QY 1735 gctagtgagctgtagaaaaatgctaaaatctaaacagatataagcaacatagattcagct 1794
DB 1561 GCTAGTGAGCTGTAGAAAAATGCTAAATCTTAATACAGATATAAGCAACATAGATTCAGCT 1620
QY 1795 aaagcagcaatcgctactgctaagacacaaaatagctgaagctcagaaaaagttcccccag 1854
DB 1621 AAAGCAGCAATCGCTACTGCTTAAGACACAATAAGCTGAAGCTCAGAAAAAGTTCCCGAC 1680
QY 1855 tctccaatcttcaagaagcggaaacaaatggtataacaggtgagaaagatctttaaaat 1914
DB 1681 TCTCCAAATCTTCAAGAAACGCGAAACAAATGGTAAATACAGGCTGAGAAAGATCTTAAAT 1740
QY 1915 atcaaacctcgagatggtctgagttccaaatccagaaactcagttggaggtcccaag 1974
DB 1741 ATCAAACTCGAGATGGTCTGATGTTCCAAATCCAGGAACCTACAGTTGGAGGCTCCCAAG 1800
QY 1975 caacaagaagtagtattggtagtattctggtttccatgctgttagatgtagtgaataat 2034
DB 1801 CAACAAGGAAGTAGTATTGGTAGTATTGCTGTTTCCATGCTTGTAGATGATGCTGAAAT 1860

```



```

/transl_table=11
/product="hypothetical protein"
/protein_id="AADI18870.1"
/db_xref="GI:4377036"
/translation="MSPLLIKPFSPPLKQDYLFDISPYTSSEITIGGSYFKLNKASL
QSPLRLRSIIIS"
gene      8654..9535
/feature="info"
/feature="CpN0732"
8654..9535
/feature="info"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AADI18871.1"
/db_xref="GI:4377037"
/translation="MKVLPPPSIPLLAHTSTAGLKNALIVEGRDIGASTVQIFITANQ
RQWRRALKREVIDEFAALKETDLSYIMSHAGYLINPGADPDVILKSKRIGTYQETL
DCITLIGSFVNFHPCALKSKSKEDMKIYVSFSQAPLFDSSPLVLVLETTAGCOT
LIGNFELGVLNOLKNOPIGVCDTCHIFAAGYDITSPQCHWEDVLEFDEVGLS
YLRAFHLNDSMPFLGANKDRHAPLGEYIGKESFKLMTDERIKKPKYLETPGGPEN
WQKEIGELKFSKNRDS"
gene      complement(9654..10283)
/feature="rs4"
/feature="CpN0733"
complement(9654..10283)
/feature="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AADI18872.1"
/db_xref="GI:4377038"
/translation="MARVCGPKNVRARFGRNIPGRSNPLLLKPKHPGOGHQMORKKK
SDYGLQLEKOKLAKCYIMKEQLVRAFKVEYIHKQNVQAFLEPERECRLDNWYRM
GFAPKIFEAQQLVAHGLHVLNVRDRRSFVFLPGMQISLKEKRSKRSQVKALESKD
RSSLSYSLDKTQFKGELLVSPEDQTEAQLPLINISVVVCEFLSHRT"
BASE COUNT 3278 a 2158 c 1983 g 3002 t
ORIGIN

```

```

Query Match      83.4%; Score 2122.2; DB 1; Length 10421;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 2138; Conservative 0; Mismatches 8; Indels 1; Gaps 1;

QY 174 catgacaaaaacattatgcttggttttagaaggattcctcaatcgcttccctaaaca 233
Db 5799 CATGACAAAAACATTATGCTTGGGTTGTAGAAAGGATTCCTCAATCGTTTGCCTAAACA 5740

QY 234 gttttttgtaaatgtagttgttcgactggaacacattcgttcccttcagaaacctccac 293
Db 5739 GTTTTTGTGAATGTAGTTGTGCTGACTGGAAACACATTCTGTTCTTCAGAAAACCTCCAC 5680

QY 294 tacagaaaaagctgctacaaacgctatgaaatacaataactgtgttgagtgagtgctgt 353
Db 5679 TACAGAAAAAGCTGCTACAAACGCTATGAAATACAAATACTGTGTTGGCAGTGGCTGT 5620

QY 354 cgaaagcatagtcaggttcttgatcaatgacagaaaagcctcttatctttatgg 413
Db 5619 CGNAAAGCATGATCAGGTTCTTGGATCAATGGACAGAAAAAGCCCTATATCTTTATGG 5560

QY 414 agctttcttaataaacoccttagcaaggctacgaagactacgtttaaataggaaagaaa 473
Db 5559 AGCTTTCTTAATAAACCTTTAGCAAAAGGCTACGAAGACTACGTTAAATGGAAAAGAAAA 5500

QY 474 cctagctgtttattgaggaacttaggggactcagaaagctggagactggctgc 533
Db 5499 CCTAGCTTGGTTATTGGAGGAACCTTTAGGGGACCTCAGAAAAAGCTGGAGCTGGTCTGC 5440

QY 534 cacagtacgttatgtagtgcgaagccttgctcagtaaatagatgtttcagggat 593
Db 5439 CACAGTACGTTATGAGTATGTCGAAGCCTTGTTCGGTTCAGAAATAGATGTTTCAGGAT 5380

QY 594 tggccgtggaattattaaagtttgggttcgcccaagcaattgctgctaactatgatcc 653

```

```

Db 5379 TGGCGTGGTAATTTATTAAGTTTGGTTCGCCCAAGCAATTCGCTGAACATATATCC 5320
QY 654 taaagaggctaaatagtttttcaaatataaaaggatttccgctctatatatgatggcat 713
Db 5319 TAAAGAGCGCTAATGGTTTTTCAAAATATAAAGGATTTTCGCTCTATATATGATGSCAT 5260
QY 714 cacagattctctcatctcagagcttatgggcttactccaaccagcaaacataaact 773
Db 5259 CACAGATCTCTATCATTCAGACTTATGGGGCTTACTCCAAACCAGCAACGATAAAT 5200
QY 774 cggcagtgattttactttccgaaagtgtgatctaggtataaatttcagcgttttaagcaa 833
Db 5199 CGGCAGTGAITTTACTTTCCGAAAGTTTGATCTAGGTATATATTTACGCGTTTTAAGTCAA 5140
QY 834 attttaataaaacttttaaaaaacaggctcgcattaaattattagtagagagctttttttta 893
Db 5139 ATTTTAATAAAATCTTTAAAAACAGGCTCGCATTAATTAATTAAGTAGAGAGCTTTTTTTTA 5080
QY 894 tttttataataaaactaaagattttttatttttttttttttttttttttttttttttttttt 953
Db 5079 TTTTATTAATAAATAAAGATTTTATTTATTTTGTGAGTTTTTATGTTTATGTTTATCTAT 5020
QY 954 tggctccaggtcctatagacgaaacagacacacccctcccgagatctttctgctcaagg 1013
Db 5019 TGGTCCAGGTCCTATAGACGAACAGACACACACCTCCCGCAGATCTTTCTGCTCAAGG 4960
QY 1014 attggagcgagtcagcaaaataaagtcgggaagctcaagaatacgaagtcggaagc 1073
Db 4959 ATTGGAGCGAGTCGAGCAAAATAAGAGTGGCGAAGCTCAAAAGATAGCAGGTGCGGAAGC 4900
QY 1074 taagcctaaagaacttaagaccgctctgttagagcgtgagagcttttggtctgctgagct 1133
Db 4899 TAAGCCTAAAGAATCTAAGACCGATTTCTGTAGCGGATGGAGCATCTTTCGCTTCAGT 4840
QY 1134 gaatgctctatgagtcgtgagataaagtcgtggttattgcttctagtaaacagctgcttc 1193
Db 4839 GAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGTAAACAGCTGCTCTTC 4780
QY 1194 tactagcagatctcagacgtggagctcaacgacagcagccacccactcgcctccctccacc 1253
Db 4779 TACTAGCAGATCTCGACAGCTGGAGCTCAACGACGACGACGACGACGACGACGACGACGAC 4720
QY 1254 cagctctgatattataaagctcaagcgaacacagcttcacgatactatttcaactcaac 1313
Db 4719 CAGCTTTTCATGATTATAAGACTCAAGCGCAAAACAGCTTACGATACATCTATCTTACCTCAAC 4660
QY 1314 atcaactgctgacatacaggctgcttggtagcctccagagctgctgctcaataataaa 1373
Db 4659 ATFCATAGCTGACATACAGGCTGCTTTGGTAGCCCTCCAGGATGCTGCTCAATATAAAA 4600
QY 1374 ggatacagcggctactgatgagaaacccgcaatcgctcgagtgaggaaactaagaatgc 1433
Db 4599 GGATACAGCGGCTACTGATGAGAAACCGCAATCGCTCGGAGTGGGAAACTAAGAAATGC 4540
QY 1434 cgaatgcaattaaagtggcgcgcaaatcacagaattagcgaaataatgcttgcgataaaca 1493
Db 4539 CGATGCAGTTAAAGTTGGCGCGCAAAATACAGAATTAGCGAAATATGCTTCGGATAACCA 4480
QY 1494 agcagctcttgactctttagttaaactgaactccttcgacctcttcagagactgctcttct 1553
Db 4479 AGCGATCTTGACTCTTTAGGTAAACTGACTTCTTCGACCTCTTACAGGCTGCTTCTCT 4420
QY 1554 ccaatctgtagcaaaacatacaaaagcagctgagcttctttaaagagatgcaagatacccc 1613
Db 4419 CCAATCTGTAGCAAAACATAACAAAGCAGCTGAGCTGCTTCTTAAAGAGATGCAAGATAACCC 4360
QY 1614 agtagtcccagggaaaaacgctgcaattgctcaatctttagttgacagacagatgcttac 1673
Db 4359 AGTAGTCCCAGGGAAAAACGCCCTGCAATTCGCTCAATCTTTAGTTGATGACAGATGCTTAC 4300
QY 1674 agcgacacagatagagaaagatggaaatgcgattggggatgcataattttgcaggagacaga 1733

```

Db	4299	AGGCACACAGATAGACAAAGATGAAATGCGATTAGGATGCATATTTTCAGGACAGAA	4240	Medical Center Dr, Rockville, MD 20850, USA
QY	1734	cgcagtcagctgtagaataatcctaatacaataataaagcaataagattcagc	1793	Location/Qualifiers
Db	4239	CGCTAGTGAGCTGTAGAAAATGCTAAATCTAATAACATATAGAACATAGATTTCAGC	4180	1. 12548
QY	1794	taagcagcaaatcgctactgactgaaagacacacaaatagctgaagctcagaaaaagtcccccga	1853	/organism="Chlamydomydia pneumoniae AR39"
Db	4179	TAAAGCAGCAATCGCTACTGCTAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCTCCGA	4120	/strain="AR39"
QY	1854	ctctcaaatcttcaagaagcgggaacaaatggttaatacagggctgagaagatcttaaaaa	1913	/db_xref="taxon:115711"
Db	4119	CTCTCCAAATCTTCAAGACAGCGAACAATAGTAAATACAGCTGAGAAAGATCTTAAAAA	4060	/note="synonym: Chlamydia pneumoniae AR39"
QY	1914	tatcaaacctgcagatggttctgtagttccaaatccaggaactcagttggaggtcccaa	1973	297. .965
Db	4059	TATCAACCTGCAGATGGTCTGATGTTCCAAATCCAGAACTACAGTTGGAGGTCCAA	4000	/gene="CP0011"
QY	1974	gaacaagaagtagtattggttagttatctggtttccatgctgtagatgtagtgcga	2033	/gene="CP0011"
Db	3999	GCAACAGGAAGTAGTATTGGTAGTATTCGTTGCTGTTCCATGCTGTTAGATGATGCTGAAA	3940	/note="similar to SP:P31218 GB:X71492 PID:296947 GB:U00096
QY	2034	taagaccgtctccatttgaatctggttgcagatgattcacaatgttcaatacga	2093	PID:1736770; identified by sequence similarity; putative"
Db	3939	TGAGACCGCTTCATTTTGTGCTGGGTTTCTGATGTTCCAAATCCAGAACTACAGTTG	3880	/codon_start=1
QY	2094	aaatctgattctcaagctgcccacagagagctcgcagcacagctagagcagcgaagc	2153	/product="uridine kinase"
Db	3879	AAATCTGATTTCTCAAGCTGCCACAGGAGCTCGCAGCACAACTAGAGCGGAAGC	3820	/db_xref="GI:7188950"
QY	2154	cgtgagatgaagtcgctgagcgtggtgagatgctcagaaagctttagaagcgc	2213	/db_xref="GI:7188951"
Db	3819	CGCTGAGATGACAGTCTGCTGACGCTGCGAGATGCTCAGAAAGCTTAGAAGCGCG	3760	/translation="MEKXYALAYYYTRVDNPHHEIALHKKFLELDVDSRIYISEQ
QY	2214	tctagtaagctggcgaacacagggcactcaatgcttggacagatcagcttcgc	2273	GINQSGTEPHAEIYMWLKERPNFSKIKRIHHEIKENIFRITVTKRKLAAALGCE
Db	3759	TCTAGTAAAGCTGGGCAACAGAGGGCACTCAATGCTTTAGGACAGATCGCTTCGC	3700	VDLSQAARKHSPQEWHEKILQENRCLILDVNNYEWKIGFDNATLPDITOTFEPEYA
QY	2274	tgtgtgtgagcagcaggag-tctctccgctgagcagaaagtctctatg	2319	EKLAQCECDPTVMYGTGICELVSPVLEKGFKEYQLDGVAYGQVGTGKW
Db	3699	TGCTGTGTGAGCGCAGGAGTCTCTCCGCTGCGAGCAAGTTCTATAG	3653	LGLKLVFDDRLAIPIDESDPVAPIAECCHQCTPSDAYNCANTCNALFLCCDCBIH
RESULT	3			OHQCGCGECSQSPRVKFDSSRGKPFRAHLCEISENSEASACLI"
AE002165	12548 bp	DNA	BCT	2238. .2867
LOCUS	Chlamydomydia pneumoniae AR39, section 2 of 94 of the complete			/gene="CP0013"
DEFINITION	genome.			2238. .2867
ACCESSION	AE002165			/gene="CP0013"
VERSION	AE002165.1			/note="similar to SP:P81288; identified by sequence
KEYWORDS				similarity; putative"
SOURCE				/codon_start=1
ORGANISM	Chlamydomydia pneumoniae AR39.			/product="ribosomal protein S4"
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomydia.			/protein_id="AAF37909.1"
AUTHORS	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,			/db_xref="GI:7188952"
	White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,			/translation="MARYCGPNRVARRFGANIFGRSRNPLLLKPPHPPQGHQMRKKK
	Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,			SDYGLQLEKOKLKACYGMKEKOLVAFKEVHKOGNVAQMFLEPRERCLDNMYRM
	Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,			GFAKTIFAAQQLVAHGHLVNGRRVDRSFLRPGMOISLAEKSKRLQSVKDALESKD
	McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			ESSLPSYISLDTKFGKGLLVSPQDQTEAQLPLFINISVCEFLSHRT"
TITLE	Genome sequences of Chlamydia trachomatis MoPn and Chlamydia			complement(2986. .3867)
	pneumoniae AR39			/gene="CP0014"
JOURNAL	Nucleic Acids Res. 28 (6), 1397-1406 (2000)			/gene="CP0014"
MEDLINE	20150255			/note="similar to SP:P12638 GB:M22591 PID:146954
PUBMED	10684935			PID:405898 GB:U00096; identified by sequence similarity;
REFERENCE	2 (bases 1 to 12548)			putative"
AUTHORS	Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,			/codon_start=1
	White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,			/product="endonuclease IV"
	Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,			/protein_id="AAF37910.1"
	Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,			/db_xref="GI:7188953"
	McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.			/translation="MKVLPPPSIPILGHAHTAGTAGLKNAIVEGRDIGASTVOIFTANQ
TITLE	Direct Submission			RQWRRLAKEVIEDFAALKETDLSYIMSHAGYLINPCADPPVILEKSRIGIYQEIL
JOURNAL	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712			DCITIGISVFVNFPGAAKSKEDCMNRIVSFSQAPLSPPLVLLTETAGQGT
				LIGNSFELGILVONLQWIPIGVCVDTCHIFAGYDITSPQWEDVLEFDEYVGLS
				YLRAFPLNDSMFPLGANKDRHAPLGEVIGKESFKFLMTDTRIKPKYLETPGPGPEN
				WQKEIGELLKFSKNRDS"

```
gene
complement(4202..4378)
/gene="CP0015"
CDS
complement(4202..4378)
/gene="CP0015"
/Note="hypothetical protein; identified by Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:17188954"
/translation="MSFLLKPPSPPLKQDYLFDISPTSSBITTIGGSYFKNLNASL
QSSTLRSLRSIIIS"
4362..6005
/gene="CP0016"
4362..6005
/gene="CP0016"
/Note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:17188955"
/translation="MSRKDNESVLARSFINLSCFTCSRTIGIFREIATAMATYFADPI
VAARWLGRTVFFLRKLLGLLEQAFIPHEFFRAQSLDRAAFFFFRFSRLKGFPI
IFTLLIEAVLVLIQYVEEGYDMILLTLLPCGIFLMMNVNAGLLHCENKFFGVG
LAPVNVNIFWVIAARHSDPRRIIGLSVALYVGFEEFWLIITVPGWVWLEAKSP
PQGHVSFVALLPGLSLTSSIFQLNLLSICLARYVHEIGPLMYSLKIVQLPIH
LFGHVFVLLPATSRVOREDERGLKMLFVLTMTSMVMIIMTAGLLIALPGLRV
LYEHLFPOSAYAVIARVLCYGASIIIPMALPIVSVLFAOROYAVPLFTIGITALA
NIVLSLWGRWLKDVSGISATSIATWVQLYFLWYSSKELPMYSLJMESIRRSIK
NVTMTLACMTLGNILITQTYVIFLNPPLPLAWSSITAOAIALFSESCIFLAFI
FGFALKLVEEDLINLASFEYWRGQGLLORHVMQDTQN"
6108..7376
/gene="CP0017"
6108..7376
/gene="CP0017"
/Note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:17188956"
/translation="MKQVYQWLASVLLALTISGYAELPLSEOKVSKHTVTTLDEVK
DYLKRGVETRRKQGVLRKAGDVRRWLYFREDIKNPSDKDKYNPLPVNRYRSEFYL
YIDYRAERNWLSKMNWTAAGGENTAAAGVDINRAFLGYRKYKNPSTRDFFMEIGRS
GLGDLFSEVQPSQNFQDGLHIYWTRELSKDYPIQYIVHGGPFVNMVTKKHAWVVEGI
LNLRLPKQFVKQSVVDWNTFVPSSTTEKATNAMYKYCVWQWLVGKHQSPWVING
OKKPLLYGAFILMNPPLAKATKTTLNGKENLAWF IGTTGLGLRKAGDMSATVRYEVEA
LSVPEDVSGIGRGLLKFQFAQIAANYDPKEANGFTNYKGFSAIYWGVTDSLSFR
AYGAVSKPANDKLGSDFTFRKFDLGIISAF"
7488..9443
/gene="CP0018"
7488..9443
/gene="CP0018"
/Note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:17188957"
/translation="MVNPIGPGPIDETERTPPADLSAOGLEASAANKSAEAOIRIAGE
AKPKESKTSVERWLSILRSVALMSLADKLGIASSNSSSTSRANDVDSSTATAPTP
PPPTFDYTKQAQITAYLTFTSTSLAQIALVLSQDAVTNKKIDPADEETIAAEW
ETKNADAVKVGQAQITELAKYSADNOALDLSGLKLTSLDQLLQALLOSVAENKAAELL
KEMQDNVPVGPPTPAQSLVDQTDATQIEKDNAINADYFAQGNASGAVENAKSN
KTSNTSDSKAAATATAKTQIAEAQKFPDPSPIQAEAOQVIAQERKLNKIPADGSDV
PNPCTTVGSGKQSGSISIRYSMLLDNAENETASTILMSGFRQMTMHMFENTPDSQVA
QBELAAQAAKAAAGDSDAAALADAQKALEALGKAGQOGGILNALGQIASAAVVA
GVPPAAASSIGSSVKQIKYTSKSGSDYKIQISAGVDAYSINDAYGRANDATRDVI
```

```
NNVSTPALTRSVPRARTEARGPEKTDQALARVIGSNRSLTGLDVYSQVSALOSVMQIIQ
SNPOANNEEIROKLTSAVTKPPOFGYPVVLNSDSTQKFIARLESFAGSRTAAEIK
ALSFTNSLFTQOVLVNIQSLYGLQ"
complement(9770..12388)
/gene="CP0019"
complement(9770..12388)
/gene="CP0019"
/Note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37915.1"
/db_xref="GI:17188958"
/translation="MSTFISQNRLLRTISGESTRIIKLDHKYSGDFPRSPVAINLEELN
SGIYALRHLMALOSENTNVAALLNPNTIIFPTSWTDYKHSRPOASSPRAPSSOTPT
DIVSAAALALVLDIGDLAEVSALEIDLGALSTISTVYRQLMASVLGTLTAQEQEK
VYSSSVPEKSLNLEHVKOEAEEIAQOEIKAVLEAGVSTSEIEAILKEYPDYIY
AADFFKEIEEPLHTYRAKVGAPIQEMNENAIQLLPPTPAITPDNVNVEVNGMTLSLI
LOADDAIKOAPALGDDQEIITLQVPLVDKTTFTKAEFDLIYTAQLPNTASLKL
YLTDRQIAEVRGKITKVYQNSIQNLSKRVVNNRSMLETOLOFOQAOQNFVTVIS
QANALNIAITNKYISAVLTSMEMVGLLCLSVMYERLADDEKATFDKSVNLYLPHI
VVGGSWVNIWAKMAAYQELAEYSLSGTAVTSQDIKAYLQTRGNFKAPRHFHNGD
OMYQFANETVFGNCLITANGAIQPDLDGGFIREAMTVNGIVEADYVSNARILNEFNTA
ATAHVLIQLOIAELQKADLDLPKASFTENRKFVAAWITSESIGLDALISMLNSQ
LPKQEAFLKPLIEEINFNNLAANLSLQITNEFTTSVYVYSSLYVQSKTGQNLF
```

Query Match	83.4%	Score 2122.2	DB 1	Length 12548
Best Local Similarity	99.6%	Pred. No. 0		
Matches 2136	Conservative 0	Mismatches 8	Indels 1	Gaps 1
Qy 174	catgacaaaaaacattatgctgtgggttagaaggattctcaatgcttgctcctaaca	233		
Db 6722	CATGACAAAAAACATATGCTGTGGTGTAGAAAGGATTCTCAATCGTTGCTAAACA	6781		
Qy 234	gtttttgtgaaatgtagtgtgtcgactggaacacacattcgcttcttcagaaacctcac	293		
Db 6782	GTTTTTTGTGAAATGTAGTGTGTGACGTGACACACATTCGTTCTTCAGAAACCTCCAC	6841		
Qy 294	tacgaaaaagcgtctacaaacgctatgaaatacaataactgtgtgttgccagtggtcgt	353		
Db 6842	TACGAAAAAGCCTGCTACAAACGCTATGAATACAAATACATCTGCTTTGGCAGTGCCTG	6901		
Qy 354	cggaaagcatagtcaggttccttgatcaatggaacagaaaaagcctctatatctttatgg	413		
Db 6902	CGGAAAGCATAGTCAGGTTCCTTGGATCAATGGACAGAAAAAGCCTCTATATCTTTATGG	6961		
Qy 414	agctttctaatgaaccccttagcaagcctacgaagactacgttaaatggaaagaaaa	473		
Db 6962	AGCTTTCCTTAATGAACCCCTTAGCAAAAGCCTACGAGACTACGTTAAATGGAAAGAAAA	7021		
Qy 474	cctagcttggtttatttgaggaaacttttaggggagactcagaaagctgagagactggctgc	533		
Db 7022	CCTAGCTTGTTTATTGAGGAACCTTAGGGGAGCTCAGAAAGCTGAGACTGCTCTGTC	7081		
Qy 534	caacagtcttatgtagtattgagaagccttgcagttccagaaaaatagattgttcaggat	593		
Db 7082	CACAGTACGTTATGATGATGTGCAAGCCCTTGTCTGGTTCCAGAAATAGATGTTTCAGGGAT	7141		
Qy 594	tggccgtggttaattattaaagtttggctgcacaaagcaaatgctgctaactatgatcc	653		
Db 7142	TGGCCGTGTTAATTTATTAAAGTTTGGTTTCGCCCAAGCAATGCTGCTAACTATGATCC	7201		
Qy 654	taaaagagcctaagttttacaaaataaaaggattttccgctctatatatgattgagcat	713		
Db 7202	TAAAGAGGCTAATGGTTTTCACAAATTAATAAGGATTTTCCGCTCTATATATGATATGGCAT	7261		
Qy 714	caacagtctctatcattcagacttatgggcttactccaacacagcaaacgataaact	773		
Db 7262	CACAGATTCCTATCATTACAGAGCTTATGGGCTTACTCTCAACACAGCAAGATAACT	7321		
Qy 774	cggcagtgatttactttccgaaagtgtgatagataatttcagogtatttaagtcac	833		

Db	7322	CGCAGTGATTTTACTTTCGCCGAAGCTTTGATCTAGGTATTAATTTTCACGGTTTAAAGTCAA	7381
Qy	834	attttaataaatcttttaaaacagcgtcgcattaataatttagtgaagacttttttttta	893
Db	7382	ATTTTAAATAAATCTTTAAAAAACAGGCTCGCATTAATTTATTAGTGAGAGCTTTTTTTTTTA	7441
Qy	894	ttttttataaataaactaaagaattttttatttttttttttttttttttttttttttttttttttttt	953
Db	7442	TTTTTTTATATATAAATCTAAAGAGATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTT	7501
Qy	954	tggctccaggttccttatagacgaacgaacgcacacctcccgagatctttctgtctcaagg	1013
Db	7502	TGGTCAGGTCCTTATAGACGAACGAACGACACACTCCCGCAGATCTTCTGCTCAAGG	7561
Qy	1014	attggaggcgagtcgagcaaaaataagagtcgggaagctcaagaatacagcaggtgcggaagc	1073
Db	7562	ATTGGAGGCGAGTGCAGCAAAATAAGAGTGCAGGAGCTCAAAAGAAATAGCAGGTGCGGAAGC	7621
Qy	1074	taagcctaagaatacttaagaccgattcttgtagagcgatggagcattcttggtttctgcagt	1133
Db	7622	TAAGCCTAAGAATCTTAAGACCGATTCTGTAGAGCGATGGAGCATCTTGCGTTCTGCAGT	7681
Qy	1134	gaatgctctcatgagtcctggcagataaagctgggtattgtctctctagttaaacagctgtcttc	1193
Db	7682	GAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTTCTTAGTAACAGCTCGTCTTC	7741
Qy	1194	tactagcagatctgcagacgtggactcaacgcagcagccagcactcagctctctccacc	1253
Db	7742	TACTAGCAGATCTGCAGACGTGGACTCAAGCAGCAGCAGCCGCACTAGCGCTCTCTCCACC	7801
Qy	1254	cacgtctgatgattataagactcaagcgcgaacagcagcttacgatactatttttaacctcaac	1313
Db	7802	CACGTTTGATGATTATTAAGACTCAAGCGCAACAGCTTACGATACTATCTTTACCTCAAC	7861
Qy	1314	atcacctagctgcacatacagctgctttgttgtaagctccagagactctgcactaataataa	1373
Db	7862	ATCACCTAGCTTGACATACAGGCTGCTTTGGTGAGCCTCAGGATGCTGTCTACATAATAAA	7921
Qy	1374	ggatacagcggcttactgataggaaacccgcaatcgctcgagtgggaaactaagaatgc	1433
Db	7922	GGATACAGCGGCTACTGATGAGGAACCCGAATCGCTCGCGAGTGGGAACATAAGATGTC	7981
Qy	1434	cgatgcgaattaaagtggcgcgcaaatcacagaattagcgaataatgcttcggataacca	1493
Db	7982	CGATGCAGTTTAAAGTTGGCGCAAAATACAGAAATACGAAATATGCTTCGGGATAACCA	8041
Qy	1494	agcgattcttgactctttaggtaaaactgactctcttcgacctcttcagagctgctctct	1553
Db	8042	AGCGATTCTTGACTCTTTAGGTAACAGTGCATTCTCTTCGACCTCTTACAGGCTGCTCTCT	8101
Qy	1554	ccaatctgtagcaaaccaatacaaacagcgtcgagctctcttaagagatgcaagataaccc	1613
Db	8102	CCAATCTGTAGCAAAACAAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCC	8161
Qy	1614	agtagtccaggggaaaaacgcttgcaattgctcaacttttagttgatcagacagatgctac	1673
Db	8162	AGTAGTCCAGGGAACCCCTTGCAATTGCTCTCAATCTTTTAGTTGATCAGACAGATGCTAC	8221
Qy	1674	agcgacacagatagagaaagatggaaaatcgatggggatgcatattttgcaggacagaa	1733
Db	8222	AGCGACACAGATAGAGAAAAGATGGAAATTCGATTTAGGATGCATATTTTTCAGGACAGAA	8281
Qy	1734	cgctagtgcagctgtagaataatgctaaatctaataacagtaataagcaacatagattcagc	1793
Db	8282	CGCTAGTGAGCTGTAGAANAATGCTTAATCTAATAACAGTATAGCAACATATAGATTGAGT	8341
Qy	1794	taaagcagcaaatcgctactgcttaagacacaaatagctgaagctcagaaaaagttccccga	1853
Db	8342	TAAAGCAGCAATCGCTACTGCTTAAGACACAAATAGCTGAACCTCAGAAAAAAGTTCCCGA	8401
Qy	1854	ctctccaattcttcaagaagcggaacaaatggtaatacagcgctgagaaaagatcttaaaa	1913
Db	8402	CTCTCCAATTTCTCAAGAAGCGGAACAAATGGTATACAGGCTGAGAAAGATCTTAAAAA	8461

Qy	1914	tatcaaacctgcagatggttctgatgttccaaatccaggaaacacagttggaggctccaa	1973
Db	8462	TATCAAACTGCAGATGGTTCTGATGTTCCAAATCCAGGAATACAGTTGGAGGCTCCAA	8521
Qy	1974	gcaacaaggaagttagtattgtagtattcggtgtttccatgcgttagatgatgcgtgaaaa	2033
Db	8522	GCAACAAGGAAGTAGTATTGGTAGTATTCGTGTTCCTATGCTGCTTAGATGATGCTGAAAA	8581
Qy	2034	tgagaccgcgttccattttagtctctgggtttctgcagatgattccatgttcaatcacgga	2093
Db	8582	TGAGACCGCTTCCATTTTGATGCTGGTTCCTGCAGATGATTCACATGTTCAATACGGA	8641
Qy	2094	aaatcctgattctcaagctcccaacagagagctcgagcaccaagctagagcagcgaagc	2153
Db	8642	AAATCCCTGATTCTCAAGCTCCCAACAGGAGCTCGCAGCACAAGCTAGACGACGAAAGC	8701
Qy	2154	cgcctggagatgacagtgctgctgcagcgcgtgcagatgctcagaaagcttttagaagcgc	2213
Db	8702	CGCTGGAGATCACAGTCTCTCGACGGCTGGCAGATGCTCAGAAAGCTTTTAAAGCGCG	8761
Qy	2214	tctaggtaaagctgggcaacacagggcattactcaatgctttggacagatgcgtcttcgc	2273
Db	8762	TCTAGGTAAAGCTGGGCAACAACAGGCGCATACATGCTTTAGGACAGATCGCTTCTGC	8821
Qy	2274	tctgtgtgaagcagcagag-tcctcccctgcagcagaattctatgg	2319
Db	8822	TCCTGTTGTGAGCGCAGGAGTTCTCCCGCTGCAGCAAGTTCTATAG	8868
RESULT	4		
AP002547/c			
LOCUS	AP002547	300550 bp	DNA BCT 08-DEC-2000
DEFINITION	Chlamydoiphila pneumoniae J138 genomic DNA, complete sequence, section 3/4.		
ACCESSION	AP002547	AB033786	AB033787
VERSION	AP002547.2	GI:101766593	BA000008
KEYWORDS			
SOURCE	Chlamydoiphila pneumoniae J138 (strain:J138) DNA.		
ORGANISM	Chlamydoiphila pneumoniae J138		
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydoiphila.		
AUTHORS	Shirai,M., Hirakawa,H., Kimoto,M., Tabuchi,M., Kishi,F., Ouchi,K., Shiba,T., Ishii,K., Hattori,M., Kuhara,S. and Nakazawa,T.		
TITLE	Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA		
JOURNAL	Nucleic Acids Res. 28 (12), 2311-2314 (2000)		
MEDLINE	20330349		
REFERENCE	2 (bases 1 to 300550)		
AUTHORS	Shirai,M.		
TITLE	Direct Submission		
JOURNAL	Submitted (04-JUL-2000) to the DDBJ/EMBL/GenBank databases. Mutsunori Shirai, Yamaguchi University School of Medicine, Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi 755-8505, Japan (E-mail:mshirai@pepo.cc.yamaguchi-u.ac.jp, Tel:81-836-22-2227, Fax:81-836-22-2415)		
COMMENT	On Sep 15, 2000 this sequence version replaced gi:6172298 gi:6172300 gi:6172396 gi:6172398 gi:8978889. AB033786-AB033787, AB033816-AB033817: Submitted (25-Oct-1999).		
FEATURES	Location/Qualifiers		
source	1. 300550		
	/organism="Chlamydoiphila pneumoniae J138"		
	/strain="J138"		
	/db_xref="taxon:138677"		
gene	/note="synonym:Chlamydia pneumoniae (strain J138)"		
	complement(21..794)		
	/gene="dapF"		
CDS	complement(21..794)		
	/gene="dapF"		
	/codon_start=1		
	/transl_table=11		
	/product="diaminopimelate-epimerase"		
	/protein_id="BAA98725.1"		

```

/db_xref="GI:8978890"
/translation="MAFISPTSTIKFYISGAGNRLLGETLVEDEVDFLCOETVRD
GFYLKPSDAQOILIFNSDGRPTMCGNLCAIAHLASQKSDISVSTDSGLYS
GYFYSWRLVMDTADWRASVHRLESRDPLPEKVCYIHTGPHAVWILPEISTDL
SILGFLRYHOTSPDCQVNVNFQILGHCOLRVYTERGVEGETAACGTGALASALV
SNSWGKESIOIHTWGELMTVSNQRGVYLOGSVTRDL"
complement(763..1338)
/gene="clpP_1"
/gene="clpP_1"
complement(763..1338)
/gene="clpP_1"
/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHKLRDIIIEKELLEARRVFFSEPVTEKSASDAIKKLWYL
ELKOPGPPIVFINSPGSDAGFAVWDQIKMLTSPVTVTGLAASMGSVLSICAAP
GRRATPHSRIMIHOPISGPIITGQATDLDIHAREILTKTARIIDVYVEATNQPDII
EKAIDRWMTANAKDFGLDGLIFSFNDL"
complement(1358..2917)
/gene="glyA"
/gene="glyA"
complement(1358..2917)
/gene="glyA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/translation="MLKVFKEFKFAIVEIFTKVAVVSLHLKLENASKKGOSLAS
TAYLAADHLNAPFSGIERIIDELKSQRSHUKMIASENYSLSVLAMGNLLTDKYC
EGSPFKRYSCCENDAIEMECVETAKELFACACQVPHSGADANLAVMAILTHKV
OGPAVKIGYTVNELTEETETLLKAEBSACVCLGSLNSGGHLTHGNVRNLVMSKLM
RCFVNDPDECPDYAIRLAKYEPKVLIAGYSSVSRRLNFAVLKQIAEDCGSLV
WYDAHFRAGLVAGGVFDEENPIPYADIVTTHKTLRGPBGGLVLAIREVESPLNK
CPLMMGGPHVIAKTAVALKEALSVDYFKYAHQVNNARLAERFISHLRLITGCT
DNHMVIDLGSIGTSKTAEDILSSVGIANNRSLPSDAIGKWDTSIRLGITPALTIL
GNMIDEEVADIIYKVLNRILRSCHVEGSKNKGELPEAIAQEARDRVRNLLRFP
LYPEIDLEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MLYLGLNQKTARKYQAHYLPILTLFPYAKSTPONKRALQFLPO
APHVILTSPSTHLFLSRMTLLSKLATKTYLCIGESTKERLLSFUGQKYVYATQ
ETAEIGIFLLQALPSSARILYPHSSLRAPVIREFLYNRETFPSPHYTVKPRKLKNI
LSKYKKIITSPSTVRAPAKIFPRPEPTYWCQGRMTLQEPKFTSSQKQVSLLETGK
SRTPS"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/translation="MASSATPGFDGTAAPSLPPATPRPNKFLALFVITAIATVWIAL
IATTAIGICIHPLCSIFLTAIFLYFISRYTCSYARNVYIALDVDPHSLQDMRS
HSPIFSDR"
4233..5302
/gene="CPj0524"
4233..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"

```

```

/translation="MSGPSRTESQSVLSVYVPRDKELAPKKQFTIAKISTLAILASL
ALGAVAGISLILVGNPVFALLITTALESVTVFVYHOMTSKYVSNQKVEQNFK
PLGKWAQENVCYSNEMQFYNNHNPKEKVAIOTDASQPOFTLTGLRVTEKQNST
GIIFNPVGTNIDINTATNLSTILSTLKDSVWOTCQREGGPAKEDPSPTEVRV
VKLPENADQTFNLNLSAAKSSILPTFLGHVCGSPKSELPPNQQETIYKQALLAENCL
KAAIESHAAIVALPLFTSVYEPPEILFKEGTFYWDNQTOAFCK3ALLDAIQNTALR
YFORSLLVILQDPFNTIESQRSEE"
5642..6406
/gene="CPj0525"
5642..6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="Ct398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MHDLSILIAQELDIKIMLRMRVKKHKEKQELAKVQSLKSDIRR
KVEKELEMLNKTQIRDGENIETISEQINKLENOAAVKKMDENALTQEMTANK
ERRSLEHQLSDLMKQAGGEDLIVLSKSLASTENSSVIERIESIKKINEEKKAL
LEORTLKHATNPELLSIYERLLNNKDRVVPVPIENRVCSGCHIVLTPQHENLVRRKD
RLIFCEHCSRILYQESQVNAQENSTAKRRRRRAV"
complement(6827..7816)
/gene="kpsF"
complement(6827..7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutO/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/translation="MPSPMISTDVCODILGKQEAVDFFQAFQPKAENQLAELKILGH
SGWFFSGVSGKCVKATVATQISLERALFFSPVDLLHGDGLVSPGDIVCLFSKS
GETQELIDTVPLHLSRAILVAITSMYFNLAALSDLVILVLSAELDPFNLIPTNST
TCOMIFDGLAMLLFHSRGLSTYGNKHPSPGQVGMKANGKVDKFPKTEVFPCHLG
DKVSELSVEFSAYCGCVCIVDPQFLMGLFTDGLRRSLASVGGVEGLSLSLEKVMTA
NPRCIEDSDIALQLMESSSPVAVLPVLDNEENRHVTGLLHMHTLAKAGLL"
complement(7843..9024)
/gene="sucB_2"
complement(7843..9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrodiploamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MIFEPFRKIGETSSGGSIVRMLKNLGDHVARDEPLIEVSTDKI
ATELPSKAGRLVRCFVNEGDEVASGDLVGLIELEETISEADDESTCPLTSCETKSEA
GSSSSVWFSPAVLSLAOREGIGLDNLQIAGTKGGRVTRDLEAVISESQOVSIPK
IFQGEVNRIPMSPLRRAIASLSKSDSEVPHASLVVDVDTDLMLNLTSGRQRFLOTH
GVKLITTSFIVQCLALRLOFPLLNGSLDGTITVMKSVNMGVAVNLNKEGVVVPVTH
NCODRGLVSAKALADLSSRARLNKLDPSEVQDGSVTITNFMTGALIGMPIIRYPEV
AILGIGTIQKRVVVRVDDDSLAIKRMVYVTLTFDHRVLDGIYGEFTLSLKNRLESVTM
G"
complement(9038..10282)
/gene="gltr"
complement(9038..10282)
/gene="gltr"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/translation="MKLWMKIFGLFVGTGLVLEDKAIFPKFIGDIFNLNLSMVVY
PLVFCSSVLTGASISDMKKLGRIGISVGLVLTALAIYIGLICFANIFSPGNCDEA
QAQSDMSAVTVIDSNTAAVFLSIIAQVFPNPNVRSFAEGNLIQIILFALFLALRL
SGERGPRVERFIDGSEITMLRMVNMINSFAPYGVGASMAWISGNHGLVQLGKFI
AYLACLFLHATLVFGGLFCGKMSFKSLSSMDAISCAVSTASSATLPVTMRCVVS
KNLGVSAFVSGVLPVIGATVNMNGTAIFQGMAAVFIQAAYNCPLSLSLILLVVTATF
SAGSAGVPGGGMITLGSVASVGLPIQGIALLAGIDRLRDIIVCTPNNILGDVAVATY
VAGGELSPYESIKQESVEET"
complement(10282..11379)
/gene="ycaH"
complement(10282..11379)

```

```

/gene="ycat"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="gi:8978900"
/translation="MKRRPSTLFLVRRVITAIISLEGILGWGLSLLSKVFAFLVA
CWNRFSTPRASTVTSVGNIVGGAGKTPTVWLAEALRLRGYSGVLSRGYKSO
SSRQKLTVDKSVHSASYGDEPLLMAEKPEGSVWPKDRRISAAARAANKFGILL
DDGLQYRLKHDVLAIVNGDPLGGRAFPKRLRDPPLRLKTVDAIIVNGGKEAG
TVVRVSNAPOLFVKPRTIASVWTHNGERIPKEALRELRVGFGGLFPQGLPQFLNIRE
EGHILKYLPLDHAATTKKELNTEQOMRQGGGLCTEKDSVKPLRUSGEVSLIP
IAKVMRLSYVNQDDTSLNLNIEQIHKNRGN"

```

Query Match		83.4%	Score 2122.2;	DB 1;	Length 300550;
Best Local Similarity		99.6%	Pred. No. 0;		
Matches 2138;		Conservative	0;	Mismatches	8;
				Indels	1;
				Gaps	1;
QY	174	catgacaaaaaacattatgcttgggtgttagaaggattctcaatcgcttgcctaaaca	233		
DB	218364	CATGACAAAAAACATTATGCTGGGTGTAGAGGATTCTCAATCGTTTGCCTAAACA	218305		
QY	234	gtttttgtgaatgttagttgtcgactggaacattcgcttcctcagaacacccac	293		
DB	218304	GTTTTTGTGAATGTAGTGTGTGCTGACTGGAACACATTCGTTCTCAGAAAACCTCCAC	218245		
QY	294	tacagaaaaagctgtacaaacgctatgaatacaaaactactgtgttgagcagtgctcgt	353		
DB	218244	TACAGAAAAGCTGTACAAACGCTATGNAATACAAATACTGTGTTGGCAGTGGCTCGT	218185		
QY	354	cggaaagcatagtcagggttccttgatcaatggacagaaaaagcctctatatctttatgg	413		
DB	218184	CGGAAAGCATAGTCAGGTTCCTTGGTCAATGGACAGAAAAAGCCTCTATATCTTTATGG	218125		
QY	414	agctttcttaagaaccttagcaaggctacgaagactacgtttaaatgaaagaaaa	473		
DB	218124	AGCTTCTTAATGAACCTTTAGCAAGGCTACGAAGACTACGTTAAATGGAAAAAGAAA	218065		
QY	474	cctagcttggtttattggaggaaactttaggggactcagaaagctggagactggtctgc	533		
DB	218064	CCTAGCTTGTTTATTGGAGGAACTTTAGGGGACTCAGAAAAGCTGGAGACTGGTCTGC	218005		
QY	534	cacagtagttatgtagtgcgaagccttgcagttccagaaaaagtagtgcctggagat	593		
DB	218004	CACAGTAGTTATGATAGTATGTCGAAGCCTTGTCCGTTCCAGAAATAGATGTTTCAGGGAT	217945		
QY	594	tggcgtggtgaattattaaagtgttgcgcgaagcaattgctgctaactatgatcc	653		
DB	217944	TGGCCGTGTAATTTATTAAGATTTTGGTTCGCCCAAGCAATTCGCTAACTATGATCC	217885		
QY	654	taaaagagcctaagttttacaaaattataaaggattttccgctctatatatgtaggcat	713		
DB	217884	TAAAGAGGCTAATGTTTTACAAATATATAAGGATTTTCCGCTCTATATATGATGGCAT	217825		
QY	714	cacagattctctatcattcagagcttaggggttactccaaacacagcaaacgataaac	773		
DB	217824	CACAGATTCTCTATCATTACAGAGCTTATGGGGCTTACTCCAAACACAGCAACGATAAAT	217765		
QY	774	cggcagtgattttactttccgaaagtgttagctaggtataatttcagcggttttaagtcac	833		
DB	217764	CGGCAGTGATTTTACTTTCCGAAAGTTTGATCTAGTATTAATTCAGCGTTTTTAAGTCNA	217705		
QY	834	attttaaaaaattctttaaaaaacagcgctgcattattattagtagagagcttttttta	893		
DB	217704	ATTTTAAATAAAATCTTTAAAAACAGGCTCGCATTAATTAATTAAGTGAAGCTTTTATTA	217645		
QY	894	tttttttaataaaaaactaaagattttattatttttttttttttttttttttttttt	953		
DB	217644	TTTTTTAATAAAAACTAAAGATTTTATTAATTTTGGTGTATTTATGGTTAAATCCCTAT	217585		
QY	954	tggctcaggtcctatagacgaacagacgcacacccctccgcagatctttctgctcaagg	1013		
DB	217584	TGGTCCAGGCTCTATAGACGAACAGACGACACCTCCCGCAGATCTTCTGCTCAAGG	217525		

QY	1014	attgagcgagtgagcaaaataaagatgcgaagactcaaaagatagcaggtgcgaagc	1073		
DB	217524	ATGAGCGGAGTGCAGCAATAAAGATGCGGAAGCTCAAAAGATAGCAGGTGCGGAAGC	217465		
QY	1074	taagctaaagaattcaagaccgattctgtagagcgatggagcattcttgcgttctgcagt	1133		
DB	217464	TAAGCTAAAGAATTAAGACCGATTCTGTAGAGCGATGGAGCATCTTTCGCTGTCAGT	217405		
QY	1134	gaatctctcatgaactctggcagataaagctgggtatttgcctctagtaaacagctcgtcttc	1193		
DB	217404	GAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGTAACAGCTGCTCTTC	217345		
QY	1194	lactagcagatctgcagcgtgtgactcaacacagcgcgcacactacgcctccctccacc	1253		
DB	217344	TACTAGCAGATCTGCAGACGTGGACTCAACGACAGCGACCGACTACGCTCTCTCCACC	217285		
QY	1254	caagctctgattattaaagactcaagcgcaaacagcttacatactatttaccctcaac	1313		
DB	217284	CACGTTTGTATGATTATAAGACTCAAGCGCAACAGCTTACGATACTATCTTTACCTCAAC	217225		
QY	1314	atcactagctgacatacaaggtgcttggtagcctccaggtgctgctcaactaataaaa	1373		
DB	217224	ATCAGTAGCTGACATACAGGCTGCTTGTGGTGAGCCTCCAGGATGCTGTCACTAATAATAA	217165		
QY	1374	ggatacagcgctactgtagaggaacccgcaatcgctcgagtgaggaaactcaagaatgc	1433		
DB	217164	GGATACAGCGGCTACTGATGAGGAACCGCAATCCGCTGGAGTGGGAAACTAAGAAATGC	217105		
QY	1434	cgatgcaattaaagttggcgcgcaaatcacagaattagcgaataatgcttcggataaaca	1493		
DB	217104	CGATGAGTTAAAGTTGGCGCAAAATACAGAAATAGCGAAATATGCTTCGATAAACCA	217045		
QY	1494	agcgaattctgactcttaggttaaactgaactccttcgacctcttacagagctgctctct	1553		
DB	217044	AGCGATTCTTGACTCTTTAGGTAAACTGACTTCTTCGACCTCTTACAGGCTGCTCTCT	216985		
QY	1554	ccaatctgtagcaacaataacaaagcagctgagctctcttaagagatgcaagaataccc	1613		
DB	216984	CCAACTGTAGCAACAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCC	216925		
QY	1614	agtagtcccagggaaaaacgcctgcaattgctcaattcttagttgtagcagacagatgctac	1673		
DB	216924	AGTAGTCCCAGGGAAAAACGCCCTGCAATTTGCTCAATCTTTAGTTGATCAGACAGATGCTAC	216865		
QY	1674	agcgacacagatagagaaagatggaaatgcgattgggagtgatatttttcagagacagaa	1733		
DB	216864	AGCGACAGATAGAGAAAGATGGAAATGCGGATTAGGATGCTATTTTTCAGGAGACAGAA	216805		
QY	1734	cgctagtggagctgtagaaaaatgctaaaatctaataacagataaagcaacatagattccagc	1793		
DB	216804	CGCTAGTGGAGCTGTAGAAAATGCTTAAATCTTAATAACAGTATAAGCAACATAGATTACG	216745		
QY	1794	taaaagcagcaatcgctactgctaaagacacaaatagctgaagctcagaaaaagttcccccga	1853		
DB	216744	TAAAGCAGCAATTCGCTACTGCTTAAGACACAAATAGCTGAAGCTCAGAAAAAGTTCCCGA	216685		
QY	1854	ctctccaattctcgaagcggaaacaaatgtaatacagcgtgagaagatcttaaaaa	1913		
DB	216684	CTCTCCAATTCTTCAAGAAAGCGGAACAATGGTAAATACAGGCTGAGAAAGATCTTAANA	216625		
QY	1914	tatcaaacctgcagatggttcttgatgtttccaaatccagaacactcagttggaggtcccaa	1973		
DB	216624	TATCAAACTGCAGATGGTTCTGTGATTCTCCAAATCCAGGAACACTACAGTTGGAGGCTCAA	216565		
QY	1974	gcaacaaggaagtagtattgtagtattcggtttccactgctgttagatgagtgcaaaa	2033		
DB	216564	GCACAAGGAAGTAGTATTGGTAGTATTTCGTTTCCATGCTGTGTAGATGATGCTGAATA	216505		
QY	2034	tgagaccctctccattttgagctggttttcgtagcagatgattcacatgttcaatcacga	2093		
DB	216504	TGAGACCGCTTCCATTTTGTGCTGGGTTTTCGTCAGATGATTACATGTTCAATACGGA	216445		

295	Qy	acagaaaaagctgctacaaacgctatgaaatacaataactggtttggcagtggtcgctc	354
121	Ddb	ACAGAAAAAGCTGCTACAAACGCTATGAAATACAAATCTGTGTTGGCAGTGGCTGTC	180
355	Qy	ggaagcatagtcaggttccttggatcaatggacgaaaaagcctctatatctttatgga	414
181	Ddb	GGAAAGCATACTCAGGTTCTTGATCAATGGACAGAAAAAGCCTCTATATCTTTATGGA	240
415	Qy	gctttcttaatagaaccttttagcaaaagctacgaagactacgttaaatgaaaaaagaac	474
241	Ddb	GCTTCTTTAATGAACCCCTTTAGCAAAAGGCTACGAAGACTACGTTAATGAAAGANAAC	300
475	Qy	ctagctctggtttattggaggaactttaggggactcagaaaaactggagactggtcgcc	534
301	Ddb	CTAGCTCTGGTTATTGGAGGAACTTTAGGGGACTCAGAAAAGCTGGAGACTGGTCTGCC	360
535	Qy	acagtcacttatgagtagtgcgaagccctgtcagttccagaataagatgtttcaggatt	594
361	Ddb	ACAGTACCTTTATGATGATGTGCAAGCCCTGTGCGTCCAGAAATAGATGTTTCAGGATTT	420
595	Qy	ggcctgtdaattattataaagttttggctgcgcccaagaacttgcgtcaactatgatcct	654
421	Ddb	GGCCGTGGTAATTTATTAAGTTTTGGTTCGCCCAAGCAATTGCTGCTAACTATGATCCT	480
655	Qy	aaagaggctaatagttttacaaaattataaaggattttccgctctctatatatgtatggcatc	714
481	Ddb	AAAGAGGCTAATGGTTTTCACAAATATAAAGGATTTTCCGCTCTATATATGTATGCGATC	540
715	Qy	acagattctctatcactcagagcttatgggcttactccaaacagcaaacagataaact	774
541	Ddb	ACAGATTCTCTATCATCTCAGAGCTTATGGGCTTAC7CCAAACAGCAACGATAACTC	600
775	Qy	ggcagtgatttactttccgaaagtttgatctagtgataatttcagc--gttttaagtc	832
601	Ddb	GGCAGTGATTTTACTTTCGAAAGTTTGATCTAGGTATATAATTTTCAGCAGGTTTTAAGTCA	660
833	Qy	aattttaataaaatcttttaaaacagagctcgcatttaattattatgagagcttttttt	892
661	Ddb	AATTTTAAATAAATCTTTAAACAGAGCTCCGATTAATATATGAGAGAGCTTTTTTTTT	720
893	Qy	attttttaataaaactaaaagatttttatttttttggagtttttka-tgggttaactcct	951
721	Ddb	ATTTTTTATATAAACAATAAAGATTTTTTATTTTATTGAGTTTTATTGGTTAATCCT	780
952	Qy	attggtccagggtctctatagaagaaacagacgacacccctccgcagatctttctgctcaa	1011
781	Ddb	ATTGGTCCAGGCTCTATAGCAAAACAGAAAGCACACCTCCGCAGATCTTTCTGCTCAA	840
1012	Qy	gaattgagcgcagtgccagcaataaagagtcggaagctcaagaaatagcaggtgcgga	1071
841	Ddb	GGATTGGAGCGGAGTGCACAAATTAAGAGTGGGAGCTCAAGAATAGCAGGTGCGGAA	900
1072	Qy	gctaaagctcaagaatactaaagaccattctgtagagcattggagcatctttgcttgcga	1131
901	Ddb	GCTAAGCCTAAAGAATCTAAGACCGATTCTGTAGAGCGATGGAGCATCTTCGCTCTGCA	960
1132	Qy	gtaaatgctctaatgagctggcagataagctgggtattgcttctagttaacagctcgtct	1191
961	Ddb	GTAAATGCTCTCATGAGTCTGCAGATAAGCTGGGTTATGTCTCTAGTAACAGCTCGTCT	1020
1192	Qy	tctactagcagatctgcagacgtggactcaacgacgacgcacacctacgctcctcca	1251
1021	Ddb	TCTACTAGCAGATCTCAGACGCTGGNACTCAACACAGCGACCGCACCTACGCCCTCTCCA	1080
1252	Qy	cccacgtctgattgattataagactcaagcgcgaacagcttaactatctttactctca	1311
1081	Ddb	CCCACGTTTGATGATTATAAGACTCAAGCGCAAAACAGCTTACGATPACTATCTTTACCTCA	1140
1312	Qy	acataactagctaacatacagactcttggtagccctccaggatgctgtcactaatata	1371
1141	Ddb	ACATPACTACTGACATACAGGCTGCTTTGGTGGACCTCCAGGATGCTGTCTACTANTATA	1200
1372	Qy	aagatcacgcgcgtactgatgaggaacccgaactcgtgcggagtggggaactaagaat	1431

Db 1201 AAGGATACAGCGCTACTGATGAGAAACCGCAATCGTGGAGTGGGAAACTAAGAA 1260
 QY 1432 gccgatgaattaaagtggcgcgaattacagaattagcaaatatgcttcgataac 1491
 Db 1261 GCCGATGCGATTAAAGTGGCGGCAAAATACAGAAATAGCGAAATATGCTTCGGATAAC 1320
 QY 1492 caagcattctgactcttttagttaaactgactctcttcgactcttacagactctctt 1551
 Db 1321 CAAGCGATCTTGACTCTTTAGTAACTGACTTCTTCGACCTCTTACAGGCTGCTCTT 1380
 QY 1552 ctccaatctgtgcgaacaatacaaaagcagctgagctctttaaagagatgaagataac 1611
 Db 1381 CTCCTAATCTGTAGCAACAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGAAGATAAC 1440
 QY 1612 ccagtagtccaggaacacgctcaattgctcaattttagttgatacaacagatgct 1671
 Db 1441 CAGTAGTCCCAGGAACGCTGCAATGTCTCAATCTTTAGTTGATCAGACAGATGCT 1500
 QY 1672 acagcgacacagatagagaaagatggaatgcgattgggagtgcatatttttcagagacag 1731
 Db 1501 ACAGCGACACAGATAGAGAAAGATGGAATGCGATTAGGATGCATATTTTCAGAGACAG 1560
 QY 1732 aacgctagtgagctgtgagaaatgctaaatctaataacagtatagaacaatagattca 1791
 Db 1561 AACGCTAGTGGAGCTGTAGAAATGCTAAATCTAATAACAGTATTAAGCAACATAGATTCA 1620
 QY 1792 gctaaagcagcaatcgctactgcttaagacacaaatagctgaagctcagaaaagttcccc 1851
 Db 1621 GCTAAGCAGCAATCGCTACTGCTTAGACACAAATAGCTGAAGCTCAGAAAAGTTCCCC 1680
 QY 1852 gactctccaattctcaagaagcgaacaaatggtaatacacagctgagaagaatcttaaa 1911
 Db 1681 GACTCTCCAATCTTCAAGAACGCGAACAATGGTAATACAGGCTGAGAAAGATCTTAA 1740
 QY 1912 aaatacaacctgcagatggttctgattgttccaaatccaggaactacagttggagctcc 1971
 Db 1741 AATATCAAAACCTGCAGATGGTTCTGATGTTCCAAATCCAGGAACATACAGTTGGAGGCTCC 1800
 QY 1972 aagcaacaggaagtagtattggttagtattcgtgttccatctgttagatgattctgaa 2031
 Db 1801 AAGCAACAGGAAGTAGTATTGGTAGTATCTGTTTCCATGCTGTTAGATGATGCTGNA 1860
 QY 2032 aatgagacgcttccattttagtctgtggtgttcgtagatgattcagatgttcaatagc 2091
 Db 1861 AATGAGACCGCTTCCATTTGATGCTGCTGGTTTCGTCAGATGATTCACATGTTCAATACG 1920
 QY 2092 gaaatcctgattctcaagctcccaacagagagctcgacagcaagctagagcagcga 2151
 Db 1921 GAAATCTTGATCTCAAGCTCCCAACAGAGGCTCGCAGCAACAAGCTAGACAGCGAAA 1980
 QY 2152 gccgctggagatgacagtctgctgcagcgtggcagatgctcagaaagctttagaagcg 2211
 Db 1981 GCCGCTGGAGATGACAGTCTGCTGCAGCGCTGGCAGATGCTCAGAAAGCTTTAGACGG 2040
 QY 2212 gctctagtgaaagctgggcaacacagggcactactcaatgcttttggacagatgctctet 2271
 Db 2041 GCTCTAGTAAAGCTGGGCAACAACAGGCGCATACTCAATGCTTTTAGGACAGATCGCTTCT 2100
 QY 2272 gctgctgttgagcgcagagagctctccgctgcagcagagttctctatgg 2319
 Db 2101 GCTGCTGTTGTGAGCGCAGGAGTCTCCCGCTGCAGCAAGTTCTATAG 2148

RESULT 6
 LOCUS AX045131
 DEFINITION Sequence 1 from Patent WO0066739.
 ACCESSION AX045131
 VERSION AX045131.1
 KEYWORDS Chlamydia pneumoniae.
 SOURCE Chlamydia pneumoniae
 ORGANISM Chlamydia pneumoniae

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia. 1 (bases 1 to 2156)
 AUTHORS Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.
 TITLE _i(chlamydia) antigens and corresponding dna fragments and uses thereof
 JOURNAL Patent: WO 0066739-A 1 09-NOV-2000;
 FEATURES Location/Qualifiers
 source 1..2156
 /organism="Chlamydia pneumoniae"
 /db_xref="taxon:83558"
 101..2056
 /note="unnamed protein product"
 /codon_start=1
 /transl_table=1
 /protein_id="CAC17329.1"
 /db_xref="GI:11343731"
 /translation="MVNPIGPIDETERTPPADLSAAGLESAANKSAEQAQRIAGAE
 AKPKESKTSVERWSILRSVAVNLSADKLGIASSNSSSTSSADVDSTTATPTP
 PPTFFDDYKTOAQAYDTIFTSTSLADIAQALVSLQDAVTNIKDTADDETTAAEW
 ETKNADAVKVGAIJTELAKYASDNOAILDSGLKTSFLLQALLOSANNKAAELL
 KEMQDNVVPKTFPAIAGSLVDQDATATQIEKGNARIDRAYFAGQNASCAVENAKSN
 NS1NIDSAKAIATATKTOIAEAKKFPDPSILQEAOMVIOAKEDLKNIKPAGSDV
 PNPETTVGSSKQGGSSIGSIRVSMLLDDAENETASILMSGFRQMIHMFENTPDSQA
 QOELAAQARAAGAAGDSDSAAALADAKALEALGKAGQOQILNLQQTASAAVSA
 GVPAAASSIGSVKQLYKTSKSTGSDYKTOISAGYDAYKSIDAYQARNADRDVI
 NVSTPALTRSPVPRARTEARPEKTDQALARVIGNSRTLGDDVSOVSALOSVMOIIO
 SNPOANNEEIRQKLTSAVTPPKQPGYPIVQLSNDSTOKFIAKLSLEAEGSRTAAEIK
 ALSFETNSLFIQQVLVNIQSLYSILQ"
 BASE COUNT 672 a 461 c 471 g 552 t
 ORIGIN
 Query Match 57.3%; Score 1459.4; DB 6; Length 2156;
 Best Local Similarity 99.5%; Pred. No. 9.4e-296;
 Matches 1474; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
 QY 840 ataaatctttaaaacagagctcgattattatttagtgagagctttttttttttttt 899
 Db 1 ATAAATCTTTAAACAGGCTCGCATTAATTATTAGTGAGAGCTTTTATTTTATTTT 60
 QY 900 ataaataactaaagattt 959
 Db 61 ATAAATAAACATAAGAGATTTTATTTATTTTATTTTATTTTATTTTATTTTATTTT 120
 QY 960 aggtcctatagacgaacagacgacacaccccccagactcttctcgaagattgga 1019
 Db 121 AGGTCTTATAGCGAAACAGAACGACACCTCCCGCAGATCTTCTGCTCAAGGATTGA 180
 QY 1020 ggcgagtgacgaaataagagtgcggaagctcaagaagatagcaggtgcggaagctaaagc 1079
 Db 181 GCGGAGTGCAGCAATAAGAGTGCAGGAGCTCAAGAATAGCAGGTGCGGAAGCTAAAGCC 240
 QY 1080 taaagaattctaagaccgattctgttagagcgtatggagactcttgcgttcgaagtaagtc 1139
 Db 241 TAAAGAAATCTAAGACCGATTCTGTAGACGATGGAGCATCTTTCGCTTCTGCAAGTATGC 300
 QY 1140 tctcatgagctgagcagataagctgggtattgcttcttagtaacagctcgctcttactag 1199
 Db 301 TCTCATGAGTCTGGCAGATTAAGCTGGGTATTGCTTCTAGTAACAGCTCGTCTTCTACTAG 360
 QY 1200 cagatctgcagactgagactcaacagcagcagccacactacactacactacactacactac 1259
 Db 361 CAGATCTGCAGACGTTGGACTCAACGACAGCGACCGACCTAGCGCTCTCCACCCACGCTT 420
 QY 1260 tgatgattataagactcaagcgaacaaagcttacgatactacttcttaactcaactacact 1319
 Db 421 TGATGATTATAGACTCAAGCGCAACAGCTTACGATACATATCTTTACTTCAACATCACT 480
 QY 1320 agctgacatacagagctgctttggtgagcctccagagatgctgctcactaataataagatac 1379
 Db 481 AGCTGACATACAGGCTGCTTTGGTGAGGCTCCAGGATGCTGCTCACTAATATATAAGGATAC 540

QY 1380 agcggtactatagagaaaccgcaatcgctgcgagtgaggaaactaagaatgccgatgc 1439
DB 541 AGCGGCTACTGATGAGGAACCGCAATCGCTGCCGAGTGGGAACTTAGAATGCCGATGC 600
QY 1440 aattaaagtctggcgcaaatatcagaattagcgaataatgtcttcgataaccacgcat 1499
DB 601 AGTTAAAGTGGCGGCAATACAGAAATAGCGAAATATGCTTCGGATACCAAGCGAT 660
QY 1500 tcttgactcttagtgaactgaactctcttcgacctcttacagactgctcttcccaatc 1559
DB 661 TCTTGACTCTTAGTAAACTGACTTCCTCGACTCTTACAGGCTGCTTCCTCCCAATC 720
QY 1560 tctacaaacaataaagcagctgagcttcttcaagagatgcaagataaccacgtagt 1619
DB 721 TGTAGCAACAATAACAAGACGCTGAGCTTCTTTAAAGAGATGCAAGATAACCCAGTAGT 780
QY 1620 cccagggaaaacgctgcaatctctttagttgtagtcagacagatgctacagcgac 1679
DB 781 CCCAGGAAACGCTGCAATTTGCTCAATCTTTAGTTGATCAGACAGATGCTACAGCGAC 840
QY 1680 acagatagagaagatggaatgcgattggggatgcataatttgcagacagaagcgttag 1739
DB 841 ACAGATAGAGAAGATGGAATGCCATTAGGGATGCATATTTTGAGGACAGAGCGTAG 900
QY 1740 tggagctgtagaaatgctaaatcctaataacagtagtaagcaacatagattcagctaaagc 1799
DB 901 TGGAGCTGTAGAATGCTTAATCTAATAACAGTATAGCAACATAGATTTCAGCTAAAGC 960
QY 1800 agcaatcgctactgctgaagacacaatagctgaagctcagaataagttcccgactctcc 1859
DB 961 AGCAATCGCTACTGCTTAAGACACAATAGCTGAAGCTCAGAAAAAGTTCCCGACTCTCC 1020
QY 1860 aattctcaagaagcgaacaatgtaatacagctgagaaagatcttaaaatatcaa 1919
DB 1021 AATCTTCAAGAGCGGAACAATGTAATACAGGCTGAGAAGATCTTAAANAATATCAA 1080
QY 1920 acctgcagatggtctctgattgttccaaatccaggaactacagttgaggtcccaagcaaca 1979
DB 1081 ACCTGCAGATGGTCTGATGTTCCAAATCCAGSAACTACAGTTGGAGGCTCCCAAGCAACA 1140
QY 1980 aggaatgattggttagttatgttccatgctgttagatgattgctgaaatgagac 2039
DB 1141 AGGAAGTAGTATTGGTAGTATTGCTGTTCCATGCTGTAGATGATGCTGAAATAGAGAC 1200
QY 2040 cgctccatttgatgtctgggttctcgtcagatgattcactgttcaatacgggaaatcc 2099
DB 1201 CGCTTCATTTTGATGCTGCGGTTTCTGTCAGATGATTACATGTTCAATACGGAAATCC 1260
QY 2100 tgattctcaagctgcccacaagagctcgcagcaacagctagagcagcaagcgcgtg 2159
DB 1261 TGATTCTCAAGCTGCCAACAGAGGCTCGCAGCACCAAGCTAGAGCAGCAAAAGCGCTGG 1320
QY 2160 agatgacagtgtgctgcagcgtgcagatgctcagaagctttagaagcgcgtctagg 2219
DB 1321 AGATGACAGTGTGCTGACGCGTGGCAGATGCTCAGAAGCTTTAGAAGCGGCTCTAG 1380
QY 2220 taaagctgggcaacaacagggcactactcaatgctttgggacagatgctctgctgctgt 2279
DB 1381 TAAAGCTGGCAACAACAGGGCATACTCAATGCTTTTAGGACAGATCGCTTCTGCTGCT 1440
QY 2280 tgtgagcgcaggag-tectcccgcctgcagcaagttctatgg 2319
DB 1441 TGTGAGCGCAGGAGTTCTCCCGCTGCAGCAAGTTCTCTATAG 1481

RESULT 7
AX045135 1456 bp DNA PAT 24-NOV-2000
LOCUS
DEFINITION Sequence 5 from Patent WO0066739.
ACCESSION AX045135
VERSION AX045135.1 GI:11343734
KEYWORDS Chlamydomophila pneumoniae.

Chlamydomophila pneumoniae
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.
REFERENCE 1 (bases 1 to 1456)
AUTHORS Murdin,A.D., Oomen,R.P., Wang,J. and Dunn,P.
TITLE (Chlamydia) antigens and corresponding dna fragments and uses thereof
JOURNAL Patent: WO 0066739-A 5 09-NOV-2000;
Aventis Pasteur Limited (CA)
FEATURES
source
1..1456
Location/Qualifiers
/organism="Chlamydomophila pneumoniae"
/db_xref="taxon:83358"
101..>1452
/notes="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAG17379.1"
/db_xref="GI:11343735"
AKPKESKTSVERSLRISAVNALSLADKLGIASNSSSSTSRSDVDTATATPTP
PPPTFDYKTOAQAYDTFTSTSLADIALVSLQDAVTNFKDRTAATDEETAAAEW
ETKNADAVKVAQITELAKVADNOALDLSGLKTSFDLLQALLQSVANNKAEELL
KEMQDNVVPKTPATAQSLVDQTDATQIEKDGNAIRDAYFAGNAGSAGVNAKSN
NSLSNDSAKAIAATATQIAEQKFPSPILQEAOMVIOAEKDLKNIKPADGSDV
PNPQTTVGSKQSGSIRVSMLLDDAENETASILMSGFRQMHMFHTNPDPDSOAA
QQLAARAAKAAAGDDSAALADAKALEALGKAGQGGGILNALGQIASAAVVVSA
G"
BASE COUNT 452 a 308 c 331 g 365 t
ORIGIN
Query Match 56.9%; Score 1447; DB 6; Length 1456;
Best Local Similarity 99.7%; Pred. No. 3.8e-293;
Matches 1450; Conservative 0; Mismatches 5; Indels 0; Gaps 0;
QY 840 ataaaattcttaaaaaacaggtcgcatattattagtgagagctttttttttttttt 899
DB 1 ATAAAACTTTAAACACAGGCTCGCATTAATTATTAGTGAGAGCTTTTTTTTATTTT 60
QY 900 ataataaactaaaagatt 959
DB 61 ATATAAATCTAAAGATTTTATTTATTTTGTGTTTATGTTAATCTATTGTTCC 120
QY 960 aggtctctatagcaaaacagacacacacctcccgagatctttctgcctcaagattgga 1019
DB 121 AGGTCCTATAGCAAAACAGAACGCACACCTCCCGCAGATCTTCTGCTCAAGATTGGA 180
QY 1020 ggcaggtcagcaataaagatgcggaagctcaagaatagcaggtgcggaagctaaagc 1079
DB 181 GCGAGTGCACAAATAAGAGTGGGAGGCTCAAGAGCTCAAGAGTACAGGTGCGGAAGCTAAGCC 240
QY 1080 taaagaatctaaagaccgattctgttagagcgatgagcatcttgcgttctgcagtgatgc 1139
DB 241 TAAAGAACTTAAGACCGATTCTGTAGAGCGATGGAGCATCTTGGCTTCTGAGTGAATGC 300
QY 1140 tctcatgagctgcgagataagctggtgattgcttcttagtaaacagctgcttctactag 1199
DB 301 TCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGTAAACAGCTGCTTCTTACTAG 360
QY 1200 cagatctgcagacgtggactcaacgacagcagcagcagcagcagcagcagcagcagcagc 1259
DB 361 CAGATCTGCAGAGTGGACTCAACGACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGT 420
QY 1260 tgatgattataagactcaagcgaacaaagcagcttgcagatactcttcaactcaaatcaact 1319
DB 421 TGATGATTATAAGACTCAAGCGCAACAGCTTACGATACTTCTTTACTCTCAACATCACT 480
QY 1320 agctcacatagcagctgcttctgtagcgtccagatgctgtcactaataataaaggtatc 1379
DB 481 AGCTGACATACAGGCTGCTTTGCTGAGCCTCCAGGATGCTGCTCACTAATAATAAGATAC 540
QY 1380 agcggctactgatgaggaacacgcgaatcgctgcgagtgagggaactaagaaatgcgcgatgc 1439

On Jun 1, 2000 this sequence version replaced gi:7190933.

Location/Qualifiers

1. .14698

/organism="Chlamydia muridarum"

/db_xref="taxon:83560"

/note="synonym: Chlamydia trachomatis MoPn"

complement(105. .1364)

/gene="TC0906"

complement(105. .1364)

/gene="TC0906"

/note="conserved hypothetical protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF39699.1"

/db_xref="GI:7190934"

/translation="MTWGLKQIFRERSRVGKQFHCYSRYFHYLPVLAAILLPGV
SWFLSEQOWRYSYFLFPVYSSGLWLFAGLRQLRAAGQQLLETKIRKLTERDEGL
KNPLETEKROKANRLKLNHDLVERLGOAREVVOAKGRVDMEEKSRKLLHEENKQ
LQMLEAARVERNEKILEQELBOELKETLAYOQLHDEVOATFVQHNNMLDKROAY I
GNLEAKVQDLMCELRNLLQLEVGAKTELPGRPNASRDVVAQVLVLEPKIVFVETFEA
ADSLALRYTRTPDSAHNYSLACKQFDDLEENLGLFIYAFPAORVLFANALFKDW
TGHGLEDLKNKGVDVLEGFPAWERDLLETESRSGKIVIKTKAFGATPFYVCVVTLL
DKGPLAEHLVGLVPAKASFPTNLSYI"

1869. .2177

/gene="TC0907"

1869. .2177

/gene="TC0907"

/note="similar to GB:D10483 SP:P02378 GB:J01683 GB:M10428
GB:X04382; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="ribosomal protein S20"

/protein_id="AAF39700.1"

/db_xref="GI:7190935"

/translation="MGEIMAPRRPSKKVGPQKPSAEKRVITSKKOLRNQSPKSRVK
TTLKKFELAVQSGDVASISAGLSVSIADKAVKRGIFKKGKADVKARASGRACPTA
"

2399. .3202

/gene="TC0908"

2399. .3202

/gene="TC0908"

/note="conserved hypothetical protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF39701.1"

/db_xref="GI:7190936"

/translation="MAAAVASSTVPAGLLSSSITALRDRDAAALRSKLGINDFLDL
LSSSFALARLAPGASWSTADQQFSTIEETLGVVGHGANDAIQTGCTGLTGLGMEFK
TOADGSFQVDTSSNKRVLFSPLSVLSKRVTLASKMLGTVMFGSPVLPVYOLGSHARG
LGLSLAGTVSSAFDAENREVCATLTTKDKCDDSTQGNSEMERERKARAGMALL
CSVPDLAQAFCEFCVETVSNAFMGVHTAFIVGICFSLSALGNVLLIAAS"

complement(3257. .5884)

/gene="TC0909"

complement(3257. .5884)

/gene="TC0909"

/note="conserved hypothetical protein; identified by Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF39702.1"

/db_xref="GI:7190937"

/translation="MSSYLLNRPNTVSGELFKIKLTNCPSCDFKNRAKPAIDIEELN
SGLYILRLSVALERAGYLGVSVINPNKIFPGGEGIVRFEQGSVLSRTISGNTVA
DIKQATPLTSTISGTLNLTIEDIPELPMQVGSISALILMETYKSKPSLNETEOT
TVFGNAYVPSDESIAKILKMEQEIQEKDILLIQAUTAGQASQALIESRLDYEEF
VYSEYFNHVEDLWKYRATIGANTQEMLEKCSALGPDIPTSLTKDNVKNANAKSVLK
AMMEAFSNVLEVEPALGSGQIEIIVAKMIPYATKTLTIDEDLHKIYQOALPSPVEM
DYLTKQEGACKGAVKAFQATQNTQAVRADVKNENIEELEIKTTSFQQAQASLES

541 ATAGAGAAAGATGGAATCGGATTAGGATGCATATTTTCAGGACAGAACGCTAGTGGA 600

1744 gctgtagaaaatgctaaatcctaataacagtagtaagcaacatagattcagctaaagcagca 1803

601 GCTGTAGAAAATGCTTAATCTAATTAACAGTATTAACACATAGATTAGCTAAAGCAGCA 660

1804 atcgctactgctaagacacaaatagctgaagctcagaaaagttccccgactctccaaatt 1863

661 ATCGCTACTGCTTAAGACACAAATAGCTGAAGCTCAGAAAAGTTCCCGGACTCTCCAATT 720

1864 cttaagaagcgggaacaaatggtataacaggtgagaagatcttaaaaatcaaacct 1923

721 CTTCAAGAACGGGACAAATGTAATACAGGCTGAGAAGATCTTAAAAATATCAAACT 780

1924 gcagatgggtctgtatgtccaaatccaggaaactacagttggaggctccaaagcaacaagga 1983

781 GCAGATGGTCTTGATGTTCCAAATCCAGGAACCTACAGTTGGAGGCTCCAAAGCAACAGGA 840

1984 agtagtattgtagtatttgcttccatgctgtagatgtagatgtagtgaagacagcct 2043

841 AGTAGTATTGGTAGTATTCTGTTTCCATGCTGTTAGATGATGCTGAAAATGAGACCGCT 900

2044 tccatttgaatctcgggttctcagatgattcattcattcaatcaggaataatcctgat 2103

901 TCCATTTTGATGCTGGGTTTCTGATGATTCATGATTCATGATTCATACGGAATCTCTGAT 960

2104 tctcaagctgcccacacagagctgcagcaacagtagtagcagcagcgaagcgcgtggagat 2163

961 TCTCAAGCTGCCAACAGAGGCTCGCAGCACAAAGCTAGACGACGAAAGCCGCTGGAGAT 1020

2164 gacagtgcgtcagcgcgtgcagatgctcagaagctttagaagcgcgtcaggtataa 2223

1021 GACAGTGCCTGCTCAGCGCTGCGAGATGCTCAGAAAGCTTTTAAAGCGGCTCTAGGTAA 1080

2224 gctggcacaacacagggcactactcaatgctttgggacagatcgcttctgctgtgtg 2283

1081 CTTGGGCAACACAGGGCATACTCATGCTTTAGGACACATCGCTTCTGCTGTTGTG 1140

2284 agccgaggag-tctctccgcgtcagcaagttctatgg 2319

1141 AGCGCAGGAGTCTCTCCCGCTGCAGCAAGTTCTATAG 1177

RESULT 9

AE002357/c

LOCUS 14698 bp DNA BCT 26-MAY-2000

DEFINITION Chlamydia muridarum, section 84 of 85 of the complete genome.

ACCESSION AE002357 AE002160

VERSION AE002357.2 GI:8163344

KEYWORDS

SOURCE Chlamydia muridarum.

ORGANISM Chlamydia muridarum.

REFERENCE 1 (bases 1 to 14698)

AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

TITLE Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39

JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)

MEDLINE 20150255

PUBMED 10684935

REFERENCE 2 (bases 1 to 14698)

AUTHORS Read,T.D., Brunham,R., Shen,C., Gill,S.R., Heidelberg,J.F., White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T., Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B., Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J., McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.

TITLE Direct Submission

JOURNAL Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA


```
QYEDIFHDSKENVIGIKKTKALFEIACVSGWLGFGGERESSILVAFBAFHGL
LFQRDLADLUSQDDQEKPEKHMNYALLFEVAAKELLDSHESFESCKNHLHQLEGL
EPLEMLCKNVFCGWK"
complement(9483. .10100)
/gene="glmu"
/nt="CT629"
complement(9483. .10100)
/gene="glmu"
/codon_start=1
/transl_table=1
/product="UDP-GlucNAc Pyrophosphorylase"
/protein_id="AAC68233.1"
/db_xref="GI:3329076"
translation="MVLSSLSFSPPEFLYPIVSOAEFVMSILPLLEKLAHSFTSGI
HGHEEGYLNKKEITIEQAGYVSGAYICGPCIIGPYQVRHGAYIRGVITSSHC
VIGHSIKNSLKHAKAAHFAYVGSFVSGVSNLGVANCFRNLGDKTIFPHHS
ERCNTKRLKGLAFGRGVSGVNTVLPNGCVASATKILPNQIY"
complement(10153. .10836)
/gene="cpxR"
/nt="CT630"
complement(10153. .10836)
/gene="cpxR"
/codon_start=1
/transl_table=1
/product="HTH Transcriptional Regulatory Protein and
Receiver Domain"
/protein_id="AAC68234.1"
/db_xref="GI:3329077"
translation="MAGPKHVLVSENWDLFFOKELLNPEYRCTIGQYKOELSD
LVCEYLSHLPREVPKSLGFSVLVLLDFEETSVDLLDRGFYLIOTIPRLKS
ALISFLSHLSVPSIRFGPNFVHLKLTVEPESVHLTPSEGLIKRLINKGO
LCRLHLEETKNAKALVARNVDVHTASLRKLKLGAYGSRIVTLRGVYLFSDGDKK
FSQDQTKLS"
BASE COUNT 3343 a 2463 c 1958 g 3190 t
ORIGIN

Query Match 10.1%; Score 257; DB 1; Length 10954;
Best Local Similarity 61.1%; Pred. No. 1e-43;
Matches 493; Conservative 0; Mismatches 270; Indels 44; Gaps 3;

QY 175 atgacaaaaacatttgctgggttagaaggattctcaatcgcttgccataaacag 234
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2892 ATGCAGAGAAAGATGCTGTGGTCTGGAGAGTATTTGAATAACTCCAGGAAT 2833

QY 235 ttttttgaaatgtagttgtgcagtggaacacattcgttcc-----ttca 282
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2832 TTGCTTGTGAAAACGAGTGTCTTGACTGGAATACGTTAACAGCAAAACGAATGATCCA 2773

QY 283 gaaacctccactacagaaaagctgctacaaacgctatgaaatacaataactgtgttg 342
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2772 GCAGACGCAAGCGCTGCACACGACTAAACCTAATACCAAGTACGATTATTAGTATGG 2713

QY 343 cagtgctcgtcggaagcatagtcaggtctcttgatcaatggacagaaagcctcta 402
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2712 CAATGTTGTTGGAGAGCACACTATGCCATGTTTAAATGGACAAACAAAATCTT 2653

QY 403 tatctttatgagcttttttaataaacctttag----- 436
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2652 TACACTTACGGAGCCTATCTCTTTAATCCATTAGCGGAATACCAAGAACTGGAACAA 2593

QY 437 ----caaggctacgaactacgttaaatgaaagaaacacacgttgatttattga 492
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2592 TCACAACTCTTACAAACCAAAATTCACAAATGTAAGGAAACCACTGCTGGTTCATCGC 2533

QY 493 ggaacttttaggggactggaagactgtagctgtctgcacacagtcagcttatgagt 552
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2532 TGCTCTCTAGGGGGTTCGAGAGCTGGAGACTGCTGCAACAGTTCGTATGATGAT 2473

QY 553 gtcgaagccttgctcagttccagaaatagatgtttcaggaggttgccgtggttaattata 612
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2472 GTTGAAGCTTTAGCGAATTCAGAAATGTATGTCGGGGTATTGTCGGGGAACCAATG 2413
```

```
QY 613 aagttttgttgcccaagcaatgctgctaactatgatactaaagagcgaatagtttt 672
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2412 AAATATGTTGCTCAAGCTATCAACAAGGATGATCGCTTAAAGAAATCTAACGCTTT 2353

QY 673 acaaatataaagatttccgctctctatatgtatgtgcatcacagattctctatcattc 732
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2352 ACTAATAAAGAGGATTCCTATCAGTTGTATGGGCTCTGACAGATTCGGTTCCTTC 2293

QY 733 agagcttatggggttactccaacacagcaaacgataaactcggcagtgattttactttc 792
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2292 CGAGCTTATGCTGCTTATTCFAAGCCTGCTAACGATAACCTTGGTAGCGACTTCACCTAT 2233

QY 793 cgaagttgctatcagtgataaatttcagcgttttaagtcacaaatttaataaaattttaa 852
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2232 CCGTAAGTATGACCTAGTGTAAATTTCTTCATCTAATCC--CTTAAGGGATAGTCTTTTA 2175

QY 853 aaacaggctcgcatataattagtagagagctttttttttttttttttttttttttttt 912
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2174 AGAGCCACACCCAAAATAACGGATGCGGCTTTTATTGCTCTTTATAATTAAGTAA 2115

QY 913 aagattttattattttttttttttttttttttttttttttttttttttttttttttt 939
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2114 GCTTTTATTTTATTTTATTTATTTATTTATTTATTTATTTATTTATTTATTTATTT 2088
```

RESULT 11

```
CTU50732 3455 bp DNA BCT 17-MAY-2001
LOCUS Chlamydia trachomatis mvin homolog, lorif2; possible membrane-bound
protein, and 76 kDa protein homolog genes, complete cds.
ACCESSION U50732
VERSION U50732.1 GI:4165180
KEYWORDS Chlamydia trachomatis.
SOURCE Chlamydia trachomatis.
ORGANISM Chlamydia trachomatis.
REFERENCE 1 (bases 1 to 3455)
AUTHORS Myers,G.S., Grinvalds,R., Booth,S., Hutton,S.I., Binks,M.,
Kemp,D.J. and Sriprakash,K.S.
TITLE Expression of two novel proteins in Chlamydia trachomatis during
natural infection
JOURNAL Microb. Pathog. 29 (2), 63-72 (2000)
MEDLINE 20368578
PubMed 10906261
REFERENCE 2 (bases 1 to 3049)
AUTHORS Myers,G.S.A. and Sriprakash,K.S.
TITLE Direct Submission
JOURNAL Submitted (05-MAR-1996) Molecular Genetics Unit, Menzies School of
Health Research, Rocklands Drive, Darwin, NT 0810, Australia
REFERENCE 3 (bases 1 to 3455)
AUTHORS Sriprakash,K.S.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-1999) Molecular Genetics, Menzies School of
Health Research, PO Box 41096, Casuarina, NT 0811, Australia
REMARK Sequence update by submitter
COMMENT On Jan 21, 1999 this sequence version replaced gi:1255183.
FEATURES
source
1..3455
/organism="Chlamydia trachomatis"
/db_xref="taxon:813"
/nt="serovar: B"
168..1778
/nt="lorif1; likely cytoplasmic membrane protein;
possible virulence protein"
/codon_start=1
/transl_table=1
/product="mvin homolog"
/protein_id="AAD08715.1"
/db_xref="GI:1255184"
/translation="MRKDDGCLVRSFLNLSGTFSSRLTGMRLREIVNATYFGADPLV
ASFVLAFTIFFPLKLLGPIGLAFIPHFELRAQNISRAFFPSRFPCYSAIL
FTLIELGLCVMCSCTGSLFDLTLLTILPLSGIFLMMYTNVSTLLHCKKFSVGL
APSVVNVSWIGTVFLARNYDPRNRIFGLAVLVIGFILEWATVLPVNMKFLGQSKEVP
```

QERDSIRALTIAPLSLGLSMGIFQLNLLCDMLARYINEVGPPLYLMSVRIQOLPVLH
FGLGVFLVLPALSRVQDHOQGYDLRLFLVAVVMTGLFVFLPGRVRL
YEHCVPEKTAHVAIVEIRGYSIIIPMALPLVSALFYARRNYKVPMLVGLIAAVVN
MYLVIGICVCKQAVLAVATSLVSPQALMLWYACAGKSLPTVYKGLMRTFFESGRTV
ITLILAAVITIGVNIHTHTYVFIPLVPTPLVSLFDQGVFAESALFLSVLFG
LAKLKTEDLMLNISFOYWKGHOSILRN
2022..2570

CDS

/note="lorf2; possible membrane-bound protein"

/codon_start=1
/transl_table=1
/protein_id="AAD08716.1"
/db_xref="GI:1255185"
/translation="MKRYFYGVFGALLACGNTNIAFAQASMSQLSVDELDL
SSGFEVTRRDGVLRLAGDVRARWYAKEDLETTQTPAKPLMPTNRYSEFNLYDY
TAANSVTKMNWVTIAGESSAAGLDINRAFLGYRFPKPNQAEVFAIGRSLGD
IFDTSFSLIVSTEFIVLTRY"
2688..3311

CDS

/note="lorf3; possible inner membrane protein; similar to
C. pneumoniae 76 kDa protein; Method: conceptual
translation supplied by author"

/codon_start=1
/transl_table=1
/product="76 kDa protein homolog"
/protein_id="AAD08717.1"
/db_xref="GI:4163181"
/translation="MKTSVVDWNTLTAKTNDPADASAAQPAKPNKYDYLVWQVLYCK
STAMPFNGQTKNLYTGAYLFNLAEPENWKMOSTTPATKTINGKENHAWFGCSLG
GVRAGDWSALVRYEYKLRPEIDVAGIGRNQKMYFAQAIKQGLDPKESGNTYK
GVSYQVWGLTDSVFRAYAAYSKPANDNLGSDTYRKYDLGLISF"
928 a 502 c 780 g 1145 t

BASE COUNT
ORIGIN

Query Match 8.6%; Score 217.6; DB 1; Length 3455;
Best Local Similarity 59.5%; Pred. No. 1.9e-35;
Matches 469; Conservative 0; Mismatches 269; Indels 50; Gaps 4;

QY 194 cttgggttagaaggattcgaatcgtttgcttaaacagtttttgaatgaatg 253
DB 2641 CTTGGGTCGTGGAAGTATTTGAATAAAATCCCGAGAAATTCGTGTGGAACAGATG 2700
QY 254 ttgcagctggaacacattcgttcc-----ttcagaacccctccactacagaaa 301
DB 2701 TTGTTGACTGGATATGTTAACAGCAAAACGAATGATCCAGCAGCGAAGCGTGCAC 2760
QY 302 aagctgtcaaaacgtctatgaatacaatactatgctgttggcagtgctcgtcggaagc 361
DB 2761 AACGAGCTAAACCTAATACCAAGTACGATTATTAGTATGCAATGTTGGTGGGAAGA 2820
QY 362 atagtcaggttccttgatcaatgacagaaaagcctctatatctttatgagctttct 421
DB 2821 GCACAGTATGCCATGTTAATTTGGAACAAACAAAATCTTACACTTACGGAGCCTATC 2880
QY 422 taatgaaccttttag-----caaggctcagaaga 451
DB 2881 TCITTAATCCATTAGCGGAATACCAGAACTGGAACAAATCAACAACTCTGCAACCA 2940
QY 452 ctacgttaaatgaaagaaacactagctgtgtttattgaggagaactttaggggactca 511
DB 2941 AAATTACAAATGTAAGGAAACCATGCTTGGTTCATCGGCTGTTCTAGGCGGGTTC 3000
QY 512 gaaagctgagactggtctgccacagctattgagatgctcgaagcctttgctcagttc 571
DB 3001 GACAGCTGGAGACTGCTGCAATAGTTCGTTATGAGTATGTAA-----CCTTCGAC 3054
QY 572 cagaatagatgtttcaggattggcgcgttggaattatttaaatgatttgggtcgcccaag 631
DB 3055 CAGAAATGATGTCGCGGGTATTGGTCGCGGAAACCAAAATGAATATGTTGTTGCTCAAG 3114
QY 632 caattgctgaactatgactcaaaagagcctaagtgttttacaattataaaggatttt 691
DB 3115 CTATCAAAACAGGATTGGATCCTTAAGAACTACGCTTTACTAATATAAGGAGTTT 3174

QY 692 ccgctctatatgtatggcatcacagattctctatcattcagagcttatgggcttact 751
DB 3175 CCTATCAGTTGTTATGGGCTGACAGATTCGGTTCTTCGAGCTTATGCTGCTATT 3234
QY 752 ccaaacccagcaaacgataaactcggcagtgatttacttctccgaaagtgtgactagta 811
DB 3235 CTAAAGCTGCTAACGATAACCTGGTAGGACTTCACCTATCGTAAGTATGACCTAGGTT 3294
QY 812 taattcagcgttttaagtcgaatttttaataaaatcttttaaaacaggtcgcattatt 871
DB 3295 TAATTCTTCATTCTAATCC--CTTAAGGGATAGTCTTTTAAGAGCAACACCAAAATA 3352
QY 872 attagtagagcgttttttttttttttttttttaataaaactaaaagattttattatttt 931
DB 3353 ACGGCTGGGCTTTTTCCTTCCTTTTATTAATAAGTAAAGGTTTCTTATTTTATTA 3412
QY 932 gaggttttt 939
DB 3413 TATTGTAT 3420

RESULT 12
AY029367

LOCUS AY029367 6849 bp DNA SYN 09-MAY-2001
DEFINITION Synthetic construct his-tagged human friend leukemia integration 1
transcription factor (FLI1) gene, complete cds.

ACCESSION AY029367
VERSION AY029367.1 GI:14017400

KEYWORDS synthetic construct.
SOURCE artificial construct.
ORGANISM Ubi.B.T.S.; Rainey, D.R. and Meredith, D.M.
REFERENCE 1. (bases 1 to 6849)
AUTHORS Ubi.B.T.S.; Rainey, D.R. and Meredith, D.M.
TITLE Direct Submission
JOURNAL Submitted (09-APR-2001) Molecular Medicine Unit, Leeds University,
Beckett St, Leeds LS9 7TF, United Kingdom

FEATURES
Location/Qualifiers
1..6849
/organism="synthetic construct"
/db_xref="taxon:32630"
/chromosome="11"
/clone="DH6"
/note="cloned into pCDNA3.1 myc his(-) expression vector;
derived from U937"
959..2398
/gene="FLI1"
959..2398
/gene="FLI1"
/note="ETS transcription factor; involved in the
pathogenesis of Ewings sarcoma following translocation
with EWS gene on chromosome 22; may have oncogenic
potential when fused with EWS gene"
/codon_start=1
/transl_table=11
/product="his-tagged human friend leukemia integration 1
transcription factor"
/protein_id="AAK50442.1"
/db_xref="GI:14017401"

gene
CDS

BASE COUNT 1632 a 1844 c 1734 g 1639 t
ORIGIN

Query Match 7.9%; Score 200.2; DB 12; Length 6849;

Best Local Similarity 94.0%; Pred. No. 8.4e-32;
Matches 219; Conservative 0; Mismatches 13; Indels 1; Gaps 1;

QY 2314 ctatgatccgagctcggtaccagcttacg-tagaacaaaaacatcatctcagaagagga 2372
|||||
Db 2299 CTTAGGAGCTACTACGGTACCAGCTTGGGCCGCAACAAAACATCTTCAGAAGAGA 2358
|||||
QY 2373 tctgaatagcgctcgacccatcatcatcatcatcatgagtttaaacgggtctccagctt 2432
|||||
Db 2359 TCTGAATAGCGCGTCGACCATCATCATCATCATCATGAGTTTAAACGGTCTCCAGCTT 2418
|||||
QY 2433 aagtttaaacctgctgatacagctcagctcagctcagctcagctcagctcagctcagctc 2492
|||||
Db 2419 AAGTTTAAACCCGCTGATCAGCTCGACTGTGCTCTTCTAGTTGCCAGCCATCTGTGTTTG 2478
|||||
QY 2493 cccctccccctgctctctgacccctggaaggtgcccactccactgctctt 2545
|||||
Db 2479 CCCCTCCCCCGTCTCTTGAACCTGGAGGTGCCACTCCACTGCTCTT 2531
|||||

RESULT 13
AXI38932
LOCUS AXI38932 5108 bp DNA PAT 31-MAY-2001
DEFINITION Sequence 51 from Patent EP1052286.
ACCESSION AXI38932
VERSION AXI38932.1 GI:14274637

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
artificial sequence.

REFERENCE
AUTHORS Morsey,M.A. and Sheppard,M.G.
TITLE Growth hormone and growth hormone releasing hormone compositions
JOURNAL Patent: EP 1052286-A 51 15-NOV-2000;
Pfizer Products Inc. (US)

FEATURES
source
1..5108
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="pGHRH1-29WTCMV construct"
BASE COUNT 1279 a 1312 c 1231 g 1286 t
ORIGIN

Query Match 6.7%; Score 171; DB 6; Length 5108;
Best Local Similarity 100.0%; Pred. No. 1.1e-25;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcggtactgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 60
|||||
Db 4723 TTGCGGTGCTGTAAACGGTGGAGGCGAGTGTAGTCTGACGACTCTGCTGCGCGC 4782
|||||
QY 61 gcgccaccagacataatagctgacagactaacagactgtctcttccatgggtctttct 120
|||||
Db 4783 GCGCCACCAGACATAATAGTGCACAGACTAACAGACTGTCTCTTTCCATGGGTCTTTCT 4842
|||||
QY 121 gcagtcacgctgtgacacgtgtgatcagatatcgcgccgctctagacc 171
|||||
Db 4843 GCAGTCACCGTGTGTCACGCTGTGATCAGATATCGCGCGCGCTCTAGACC 4893
|||||

RESULT 14
AXI38933
LOCUS AXI38933 5108 bp DNA PAT 30-MAY-2001
DEFINITION Sequence 52 from Patent EP1052286.
ACCESSION AXI38933
VERSION AXI38933.1 GI:14274638

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 5108)
AUTHORS Morsey,M.A. and Sheppard,M.G.

TITLE Growth hormone and growth hormone releasing hormone compositions
JOURNAL Patent: EP 1052286-A 52 15-NOV-2000;
Pfizer Products Inc. (US)

FEATURES
source
1..5108
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="pGHRH1-29WTCMV construct"
BASE COUNT 1279 a 1313 c 1230 g 1286 t
ORIGIN

Query Match 6.7%; Score 171; DB 6; Length 5108;
Best Local Similarity 100.0%; Pred. No. 1.1e-25;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcggtactgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 60
|||||
Db 4723 TTGCGGTGCTGTAAACGGTGGAGGCGAGTGTAGTCTGACGACTCTGCTGCGCGC 4782
|||||
QY 61 gcgccaccagacataatagctgacagactaacagactgtctcttccatgggtctttct 120
|||||
Db 4783 GCGCCACCAGACATAATAGTGCACAGACTAACAGACTGTCTCTTTCCATGGGTCTTTCT 4842
|||||
QY 121 gcagtcacgctgtgacacgtgtgatcagatatcgcgccgctctagacc 171
|||||
Db 4843 GCAGTCACCGTGTGTCACGCTGTGATCAGATATCGCGCGCGCTCTAGACC 4893
|||||

RESULT 15
AXI38936
LOCUS AXI38936 5111 bp DNA PAT 30-MAY-2001
DEFINITION Sequence 55 from Patent EP1052286.
ACCESSION AXI38936
VERSION AXI38936.1 GI:14274641

KEYWORDS
SOURCE synthetic construct.
ORGANISM synthetic construct.
artificial sequence.
REFERENCE 1 (bases 1 to 5111)
AUTHORS Morsey,M.A. and Sheppard,M.G.

TITLE Growth hormone and growth hormone releasing hormone compositions
JOURNAL Patent: EP 1052286-A 55 15-NOV-2000;
Pfizer Products Inc. (US)

FEATURES
source
1..5111
/organism="synthetic construct"
/db_xref="taxon:32630"
/note="pGHRH1-29Yalal522CMV construct"
BASE COUNT 1280 a 1314 c 1230 g 1287 t
ORIGIN

Query Match 6.7%; Score 171; DB 6; Length 5111;
Best Local Similarity 100.0%; Pred. No. 1.1e-25;
Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcggtactgttaacggtgagggcagtgtagtctgagcagtagctgttgcgcgcgc 60
|||||
Db 4723 TTGCGGTGCTGTAAACGGTGGAGGCGAGTGTAGTCTGACGACTCTGCTGCGCGC 4782
|||||
QY 61 gcgccaccagacataatagctgacagactaacagactgtctcttccatgggtctttct 120
|||||
Db 4783 GCGCCACCAGACATAATAGTGCACAGACTAACAGACTGTCTCTTTCCATGGGTCTTTCT 4842
|||||
QY 121 gcagtcacgctgtgacacgtgtgatcagatatcgcgccgctctagacc 171
|||||
Db 4843 GCAGTCACCGTGTGTCACGCTGTGATCAGATATCGCGCGCGCTCTAGACC 4893
|||||

Search completed: February 7, 2002, 19:22:17
Job time: 12851 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 685.35 Seconds

(without alignments)
3183.618 Million cell updates/sec

Title: US-09-391-606-1

Perfect score: 2545

Sequence: 1 ttgcggtgtgttaacggtg.....tgccactccactgtccttt 2545

Scoring table:

IDENTITY_NUC

Gap 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- N_Geneseq_1101:*
- 1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT:*
 - 2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT:*
 - 3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT:*
 - 4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT:*
 - 5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT:*
 - 6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT:*
 - 7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT:*
 - 8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT:*
 - 9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT:*
 - 10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT:*
 - 11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT:*
 - 12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT:*
 - 13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT:*
 - 14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:*
 - 15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT:*
 - 16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:*
 - 17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:*
 - 18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:*
 - 19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:*
 - 20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:*
 - 21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:*
 - 22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2225.2	87.4	2238	21	AAD02066
2	2111.2	83.0	1230025	20	AA91990
3	1459.4	57.3	2156	21	AAD02063
4	1447	56.9	1456	21	AAD02065
5	1155.4	45.4	1852	21	AAD02064
6	832.8	32.7	1550	21	AA28411
7	179	7.0	1909	21	AAA27110
8	171	6.7	4912	20	AA08423
9	171	6.7	4912	21	AAA71402
10	171	6.7	5108	22	AAC86258
11	171	6.7	5108	22	AAC86259

12	171	6.7	5111	22	AAC86262	CMV IE promoter eh
13	171	6.7	5185	22	AAC86254	PGHRH-4 construct.
14	171	6.7	5188	22	AAC86266	Plasmid GHRH1-44YW
15	171	6.7	5254	22	AAC86267	Plasmid GHRH1-44W
16	171	6.7	7285	20	AA89797	DNA of pVR 1012-GP
17	171	6.7	7285	20	AA89797	Plasmid pVR 1012-GP
18	169.4	6.7	7272	20	AA89797	Plasmid pVR 1012-GP
19	169.4	6.6	5322	20	AA89797	Cytokine-encoding
20	169.4	6.6	5322	20	AA89797	Plasmid VR4151
21	167.8	6.6	5322	21	AA89797	DNA of pVR 1012-GP
22	166.4	6.5	7073	20	AA89797	DNA of pVR 1012-GP
23	166.4	6.5	7073	20	AA89797	Plasmid pVR 1012-GP
24	163.4	6.4	7001	20	AA89797	DNA of pVR 1012-GP
25	163.4	6.4	7003	20	AA89797	Plasmid pVR 1012-GP
26	156.4	6.1	5480	20	AA89797	Cytokine-encoding
27	149.2	5.9	4928	17	AA89797	Plasmid encoding h
28	146.8	5.8	150	19	AA89797	Part of the gene e
29	145	5.7	7475	20	AA89797	EP-892047 Seq ID 3
30	145	5.7	8192	20	AA89797	EP-892047 Seq ID 3
31	144.8	5.7	5974	21	AA89797	Nucleotide sequenc
32	142	5.6	3125	15	AA89797	HIV virus-1(LAT) g
33	142	5.6	7108	20	AA89797	EP-892047 Seq ID 3
34	139.8	5.5	4276	22	AA89797	Plasmid pCMVII. C
35	139.8	5.5	4282	22	AA89797	pCMV-II nucleic ac
36	139.8	5.5	4328	19	AA89797	Vector plasmid CMV
37	139.8	5.5	4328	20	AA89797	Vector CMVkm2. Ch
38	139.8	5.5	4818	19	AA89797	Vector plasmid CMV
39	139.8	5.5	5107	19	AA89797	Vector plasmid CMV
40	139.8	5.5	9600	19	AA89797	Vector plasmid pCM
41	139.2	5.5	5459	22	AA89797	Plasmid pCMVII opt
42	139.2	5.5	5882	22	AA89797	Plasmid pCMV-II-E2
43	139.2	5.5	7731	20	AA89797	pgL105 F19 chimera
44	139.2	5.5	7731	20	AA89797	pgL105 human resh
45	139.2	5.5	8068	20	AA89797	pKN100 F19 chimeri

ALIGNMENTS

RESULT	1
AAD02066	
ID	AAD02066 standard; DNA; 2238 BP.
XX	
AC	AAD02066;
DT	26-MAR-2001 (first entry)
DE	C. pneumoniae 76 kda protein truncation mutant fusion gene.

KW	76 kda protein; bactericidal; diagnosis; prevention;
KW	pneumonia; upper respiratory tract disease; sinusitis;
KW	acute respiratory disease; cough; sore throat; hoarseness; fever;
KW	vaccine; immunisation; treatment; truncation mutant; fusion gene;
KW	ds.
OS	Chlamydia pneumoniae.
OS	Synthetic.
XX	
FT	Key
FT	misc_feature
FT	Location/Qualifiers
FT	1..665
FT	/*tag= a
FT	/note= "This part of the sequence is unrelated to
FT	C. pneumoniae 76 kda gene"
FT	766..2238
FT	/*tag= b
FT	/product= "Truncated Chlamydia pneumoniae
FT	76kDa protein"
FT	2122..2238
FT	/*tag= c
FT	/note= "This part of the sequence is unrelated to
FT	C. pneumoniae 76 kda gene"
PN	W0200066739-A2.

XX	09-NOV-2000.
PD	XX
PF	XX
PE	03-MAY-2000; 2000WO-CA00511.
XX	XX
PR	03-MAY-1999; 99US-0132270.
PR	30-JUN-1999; 99US-0141276.
XX	(AVET) AVENTIS PASTEUR LTD.
PA	XX
XX	Murdin AD, Oomen RP, Wang J, Dunn P;
PI	WPI: 2000-687542/67.
DR	P-PSTB; AA711957.
XX	XX
PT	Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
PT	useful for vaccinating against Chlamydia infections -
PS	Claim 32; Fig 3; 11pp; English.
XX	XX
CC	The present sequence is a DNA coding for a fusion protein comprising a
CC	C truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
CC	residues. C. pneumoniae 76 kDa protein is used in the
CC	diagnosis, prevention and treatment of C. pneumoniae
CC	infections (e.g. pneumonia, upper respiratory
CC	tract disease, bronchitis, sinusitis and acute respiratory
CC	disease such as cough, sore throat, hoarseness, fever; and
CC	abnormal chest sounds on auscultation). C. pneumoniae sequence
CC	is also used as vaccines for immunising humans against diseases
CC	caused by C. pneumoniae.
XX	XX
SQ	Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;
Query Match 87.4%; Score 2225.2; DB 21; Length 2238;	
Best Local Similarity 99.6%; Pred. No. 0;	
Matches 2230; Conservative 0; Mismatches 8; Indels 0; Gaps	
QY	175 atgacaaaaaacattatgtcttgggttgtagaaggattcctaactgtttgcctaaacag 234
DB	1 atcacaaaaaacattatgtcttgggttgtagaaggattcctaactgtttgcctaaacag 60
QY	235 ttttttgtgaatgtagtgttcgcactggaaacacattgttccttcagaacctccact 294
DB	61 ttttttgtgaatgtagtgttcgcactggaaacacattgttccttcagaacctccact 120
QY	295 acagaaaagctcgtcaaacgctatgaaatacaaatactgtgtttgccagtggctgctc 354
DB	121 acgaaaaagctcgtcaaacgctatgaaatacaaatactgtgtttgccagtggctgctc 180
QY	355 ggaagcatagtcaggttcccttggatccaatggacagaaaaaacctctatatotttattgga 414
DB	181 ggaagcatagtcaggttcccttggatccaatggacagaaaaaacctctatatotttattgga 240
QY	415 gctttcttaatgaaccctttagcaagcctacgaagactcacgtttaattggaaaagaaac 474
DB	241 gctttcttaatgaacccttttagcaagcctacgaagactcacgtttaattggaaaagaaac 300
QY	475 ctagcttggtttatttgagaaacttttaggggactcagaaaagctggagactggctgccc 534
DB	301 ctagcttggtttatttgagaaacttttaggggactcagaaaagctggagactggctgccc 360
QY	535 acagtacggttatgagtatgctgaagccttgtcagttccagaataatagatgttttcaggatt 594
DB	361 acagtacggttatgagtatgctgaagccttgtcagttccagaataatagatgttttcaggatt 420
QY	595 gcccgctggtaatttattaagttttgtgtccgcccaagcaattgctgctaactatgatcct 654
DB	421 gcccgctggtaatttattaagttttgtgtccgcccaagcaattgctgctaactatgatcct 480
QY	655 aaagagcctaagtttttaacaattatabaaggattttccgcgtctatatatgtatggcattc 714
DB	481 aaagagcctaagtttttaacaattatabaaggattttccgcgtctatatatgtatggcattc 540

Qy 1795 aaagcagcaatcgctactgctaagaacacaaatagctgaagctcagaaaaagttccccgac 1854
 Db 1621 aaagcagcaatcgctactgctaagaacacaaatagctgaagctcagaaaaagttccccgac 1680
 Qy 1855 tctcaaatcttcaagaagcggaacaaatgtaatacacaggctggaagatcttataaaat 1914
 Db 1681 tctcaaatcttcaagaagcggaacaaatgtaatacacaggctggaagatcttataaaat 1740
 Qy 1915 atcaaacctcgagatggttctgatttccaaatccagaaactacagttggaggctccaaag 1974
 Db 1741 atcaaacctcgagatggttctgatttccaaatccagaaactacagttggaggctccaaag 1800
 Qy 1975 caacaaggaagtattgtagtattcgtgttccatgctgttagatgagctgaaaaat 2034
 Db 1801 caacaaggaagtattgtagtattcgtgttccatgctgttagatgagctgaaaaat 1860
 Qy 2035 gagaccgctccattttagctgtcgtggttctggttcagatgattcacatgttcaatacggaa 2094
 Db 1861 gagaccgctccattttagctgtcgtggttctggttcagatgattcacatgttcaatacggaa 1920
 Qy 2095 aatcctgattctcaagctgccaaacagagctcgacacaaagctagagcagcgaagccc 2154
 Db 1921 aatcctgattctcaagctgccaaacagagctcgacacaaagctagagcagcgaagccc 1980
 Qy 2155 gctgagatgacagctgctgacgctgacgctgacgctgacgctgacgctgacgctgac 2214
 Db 1981 gctgagatgacagctgctgacgctgacgctgacgctgacgctgacgctgacgctgac 2040
 Qy 2215 ctaggtaagctgggcaacaacagagctacatactcaatgcttggagacagatcgcttctgct 2274
 Db 2041 ctaggtaagctgggcaacaacagagctacatactcaatgcttggagacagatcgcttctgct 2100
 Qy 2275 gctgttggagcagagagctccctcctgagcagagcttctatgagctcgagctcggttac 2334
 Db 2101 gctgttggagcagagagctccctcctgagcagagcttctatgagctcgagctcggttac 2160
 Qy 2335 caagcttcagtagaacaacaaactatctcagaagagctgtaagtagcgcgtcgaccat 2394
 Db 2161 caagcttcagtagaacaacaaactatctcagaagagctgtaagtagcgcgtcgaccat 2220
 Qy 2395 catcatcatcatcattga 2412
 Db 2221 catcatcatcatcattga 2238

RESULT 2
 AAX91990/c
 ID AAX91990 standard; DNA; 1230025 BP.
 XX AAX91990;
 AC AAX91990;
 DT 13-SEP-1999 (first entry)
 XX
 Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 DE
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-1B01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PF Griffiths R;

XX WPI; 1999-357842/30.
 DR
 XX
 PT Genome sequence of Chlamydia pneumoniae
 PS
 PS Claim 1; Page 291-611; 1912pp; English.
 XX
 CC The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AAY34584-V35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AAY34584-V35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 83.0%; Score 2111.2; DB 20; Length 1230025;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 2138; Conservative 0; Mismatches 8; Indels 2; Gaps 2;
 Qy 174 catgacaaaaaacattatgcttgggtttagaaggattctcaatcgcttgccttaaac 233
 Db 829364 CATGACAAAAAACATTATGCTTGGTGTAGAGGATTCTCAATCGTTTCCTTAACA 829305
 Qy 234 gttttttgtaaatgtagtggctgactggaacacattcgctccttcagaaacctccac 293
 Db 829304 GTTTTTTGTGAATAGTGTGTGCTGCTGGAACACATTCGTTCTTTCAGAAACCTCCAC 829245
 Qy 294 Lacagaaaaagctgtcaaacgctatgaatacaaaactgctgttggcagctgctgct 353
 Db 829244 TACAGAAAAAGCTGCTACAAACGCTATGAATACAAATCTGTTGTGGCAGTGGCTCGT 829185
 Qy 354 cggaaagcatagctaggttctcttgatcaatggacagaaaaagcctctatcttctatgg 413
 Db 829184 CGGAAGCATAGTCAGGTTCTTGGATCAATGGACAGAAAAAGCCTATATCTTTATGG 829125
 Qy 414 agctttcttaatagaaccttttagcaaaagctacgaagactacgttaaatggaagaagaaa 473
 Db 829124 AGCTTTCTTAATGAACCCCTTTAGCAAGGCTACGAAGACTACGTTAAATGGAAGAAA 829065
 Qy 474 cctagcttgggttattgaggaacttttaggggactcagaaagctgagactgctgctgc 533
 Db 829064 CCTAGCTTGGTTATTGGAGGAACTTTAGGGGACTCAGAAAAAGCTGGAGACTGGTCTGC 829005
 Qy 534 cacagtacgttatgagtagtgcgaagccttgcagttccagaaaatagatggtttcaggat 593
 Db 829004 CACAGTACGTTATGAGTATGTCGAAGCCTTGTGGTTCAGAAATAGATGTTTTCAGGAT 828945
 Qy 594 tggccgtgtaattattataaagtttggctggcccaagaactgctgtaactatgatcc 653
 Db 828944 TGGCCGTGTAATTTATTAAAGTTTGGTTCGCCCAAGCAATTTGCTTAACATATGATCC 828885
 Qy 654 taagagggctaatagttttacaaattataaaggattttccgctctatataatgtagcat 713
 Db 828884 TAAAGAGCTAATGCTTTTACAAATATAAAGGATTTTCCGCTCTATATATGATGTCAT 828825
 Qy 714 cacagattctctatcattcagagcttattgggcttactcccaaacagcaacgataaact 773
 Db 828824 CACAGATTCTCTATTCATTTCAGAGCTTATGGGGCTTACTCCAAACAGCAACGATAAACT 828765
 Qy 774 cggcagtagattttacttccgaagatttgaactagtagtaataattcagcgttttaagtc 833
 Db 828764 CGGCAGTGTATTTACTTTCCGAAAGTTTGTAGTATGATTAATTTTTCAGCGTTTAA 828705
 Qy 834 attttaataaaatctttaaaacagcctcgacttaattattagtagagcgtttttttta 893
 Db 828704 ATTTTAAATAAATCTTTAAACACAGCTCGCATTAATTAATTAATTAATTAATTA 828645

Db	827564	GCACCAAGGAAGTAGTATTGGTAGTATTCGTGTTTCCATGCTGTAGATGATGCTGAAAA	827505
QY	2034	tsagaccgcttcacatttgatgctcggggttcgtcagatgattcacatgttcaatacagga	2093
Db	827504	TGAGACCGCTTCATTTTTCATGCTGTCGGGTTTCGTTCAGATGATTCACATGTTCAATACGGA	827445
QY	2094	aaatcctgattctcaagctgcccaacagagagctcgacacaaagctagacagcgaaagc	2153
Db	827444	AAATCTGTATTCTCAAGCTGCCCAACAGCAGCTCGCAGCACAAAGCTAGACGACGCAAGC	827385
QY	2154	cgctggagatgacagtgcctgcagcgcctggcagatgctcaaaaagctttagaagcggc	2213
Db	827384	CGCTGGAGATGACAGTGCCTGCTGCACGCGTGGCAGATGCTCAGAAAGCTTTAGAAAGCGCG	827325
QY	2214	tctagtgaaagctggcacaacacagggcactcaatgctttggcacagatcgcttctgc	2273
Db	827324	TCFAGGTAAGCTGGCGACACACAGGGCATACTCAATGCTTTAGACAGATCGCTTCGC	827265
QY	2274	tgctg-ttgtgagcgcaggag-tcctcccgctgcagcaagttctatgg	2319
Db	827364	TGCTGTTTGTGAGCGCAGGAGTTCCCTCCGCTGCAGCAAGTTCTATAG	827217
RESULT 3			
AAD02063			
ID	AAD02063 standard; DNA; 2156 BP.		
AC	AAD02063;		
XX			
XX	26-MAR-2001 (first entry)		
DT			
XX			
DE	Chlamydia pneumoniae 76 kDa full-length protein encoding DNA.		
XX			
KW	76 kDa protein; bactericidal; diagnosis; prevention; treatment;		
KW	pneumonia; upper respiratory tract disease; bronchitis; sinusitis;		
KW	acute respiratory disease; cough; sore throat; hoarseness; fever;		
KW	vaccine; immunisation; ds.		
XX			
OS	Chlamydia pneumoniae.		
XX			
FH	Key Location/Qualifiers		
FT	CDS 101..2056		
FT	/tag= a		
FT	/product= "Chlamydia pneumoniae 76kDa protein"		
XX			
PN	WO200066739-A2.		
XX			
PD	09-NOV-2000.		
XX			
PF	03-MAY-2000; 2000WO-CA00511.		
XX			
PR	03-MAY-1999; 99US-0132270.		
PR	30-JUN-1999; 99US-0141276.		
XX			
XX	(AVET) AVENTIS PASTEUR LTD.		
XX			
PI	Murdin AD, Oomen RP, Wang J, Dunn P;		
XX			
DR	WPI; 2000-687542/67.		
DR	P-PSDB; AAY71954.		
XX			
PT	Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,		
PT	useful for vaccinating against Chlamydia infections -		
XX			
PS	Claim 2a; Fig 1; 1l2pp; English.		
XX			
CC	The present sequence is a DNA coding for Chlamydia pneumoniae		
CC	76 kDa protein. C. pneumoniae 76 kDa protein		
CC	is used in the diagnosis, prevention and treatment		
CC	of C. pneumoniae infections (e.g. pneumonia, upper respiratory		
CC	tract disease, bronchitis, sinusitis and acute respiratory		
CC	disease such as cough, sore throat, hoarseness, fever; and		


```

FT CDS 101..1369
XX /*tag= a
XX WO200024901-A1.
XX 04-MAY-2000.
XX 28-OCT-1999; 99WO-GB03565.
XX 28-OCT-1998; 98US-0106037.
XX 20-SEP-1999; 99US-0154658.
XX 26-OCT-1999; 99US-0427501.
XX (CONN-) CONNAUGHT LAB LTD.
XX Murdin AD, Oomen RP, Dunn PL;
XX WPI; 2000-350742/30.
XX P-PSDB; AAY92716.
XX Isolated polynucleotide encoding a Chlamydia polypeptide useful to
XX treat, diagnose and prevent disease caused by Chlamydia infection
XX Claim 1; Fig 1A-C; 88pp; English.
XX This sequence encodes lorf2 protein of a strain of Chlamydia pneumoniae.
XX Comparison of this sequence as to the recently published genome sequence
XX of C. pneumoniae reveals that the sequence actually contains at least two
XX open reading frames, a first one in the 5' portion and a second one in
XX the 3' portion of the sequence. Despite the presence of the stop codon at
XX the end of this sequence, C. pneumoniae does make a 76 kDa product. It
XX appears possible that C. pneumoniae is able to read through this stop
XX codon and produce a full-length product terminated by the stop codon at
XX the end of the second open reading frame. There is at least one in-frame
XX ATG upstream of the start codon. This suggests that the first open
XX reading frame may form part of one or more larger open reading frames.
XX The lorf2 protein or DNA can be used as a vaccine for humans to treat or
XX prevent disease caused by Chlamydia infection. The sequences or an
XX antibody to lorf2 can be used to diagnose a Chlamydia infection.
XX Sequence 1550 BP: 472 A; 268 C; 317 G; 493 T; 0 other;

Query Match 32.7%; Score 832.8; DB 21; Length 1550;
Best Local Similarity 99.8%; Pred No. 1.2e-186;
Matches: 834; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 174 catgacaaaaaacattatgcttggtgtgtagaaggattctcaatgcttgctctaaaca 233
DB 715 catgacaaaaaacattatgcttggtgtgtagaaggattctcaatgcttgctctaaaca 774
QY 234 gtttttggaaatgtagtggctgactggaacacattcgtcttcagaaacctccac 293
DB 775 gtttttggaaatgtagtggctgactggaacacattcgtcttcagaaacctccac 834
QY 294 tacagaaaaagctgtcacaaagcgtatgaatacacaatactgtgttgagctgctcgt 353
DB 835 tacagaaaaagctgtcacaaagcgtatgaatacacaatactgtgttgagctgctcgt 894
QY 354 cggaaagcatagtcaggttccttgatcaatggacagaaaaagcctctatatctttatgg 413
DB 895 cggaaagcatagtcaggttccttgatcaatggacagaaaaagcctctatatctttatgg 954
QY 414 agctttcttaagaaacctttgacaaagcgtacgaagctcgtttaaatggaaagaaaa 473
DB 955 agctttcttaagaaacctttgacaaagcgtacgaagctcgtttaaatggaaagaaaa 1014
QY 474 cctagcttggtttattggaggaaactttaggggactcagaaagctggagactggtctgc 533
DB 1015 cctagcttggtttattggaggaaactttaggggactcagaaagctggagactggtctgc 1074
QY 534 cacagtcagttatgtagtgcgaagccttgctcagttccagaaataagattttcaggat 593
DB 1075 cacagtcagttatgtagtgcgaagccttgctcagaaataagattttcaggat 1134
QY 594 tggcgtggtaattattataaagtttggctgcgcccagcaattgctgctaaactatgac 653
DB 1135 tggcgtggtaattattataaagtttggctgcgcccagcaattgctgctaaactatgac 1194
QY 654 taaagaggctaaatagttttacaaattataaaggattttccgctctatatgtatgcat 713
DB 1195 taaagaggctaaatagttttacaaattataaaggattttccgctctatatgtatgcat 1254
QY 714 cacagattctctatcatcagagcttatggggttactccaaacagcaacagataaact 773
DB 1255 cacagattctctatcatcagagcttatggggttactccaaacagcaacagataaact 1314
QY 774 cggcagtgattttactttccgaaagtttgcattaggtatatttcagcgttttaagcaa 833
DB 1315 cggcagtgattttactttccgaaagtttgcattaggtatatttcagcgttttaagcaa 1374
QY 834 attttaataaattctttaaaacaggtcgcatttaattattagtgagagcgtttttttta 893
DB 1375 attttaataaattctttaaaacaggtcgcatttaattattagtgagagcgtttttttta 1434
QY 894 tttttataataaactaaagattttttatttttttttttttttttttttttttttttttt 953
DB 1435 tttttataataaactaaagattttttatttttttttttttttttttttttttttttttt 1494
QY 954 tggctcaggtcctatagacaaacagacacacaccccccagatctttctgctc 1009
DB 1495 tggctcaggtcctatagacaaacagacacacaccccccagatctttctgctc 1550

RESULT 7
AA27110
ID AA27110 standard; cDNA; 1909 BP.
AC AA27110;
XX 04-AUG-2000 (first entry)
XX Rhesus monkey melanocortin-3-receptor cDNA.
XX Rhesus monkey; melanocortin-3 receptor protein; MC-3R;
XX G-protein coupled receptor; GPCR; obesity; ss.
XX Macaca mulatta.
XX Key Location/Qualifiers
XX CDS 148..1119
XX /*tag= a
XX /product= MC-3R protein
XX WO200027862-A1.
XX 18-MAY-2000.
XX 05-NOV-1999; 99WO-US25747.
XX 09-NOV-1998; 98US-0107725.
XX (MERI ) MERCK & CO INC.
XX Fong TM, Van Der Ploeg LHT, Huang RC;
XX WPI; 2000-387404/33.
XX P-PSDB; AAY94427.
XX New DNA encoding Rhesus monkey melanocortin 3 receptor protein,
XX recombinant vectors and host cells, useful in methods for identifying
XX selective agonists and antagonists -
XX Claim 1; Fig 1; 54pp; English.
XX A random primed probe from human melanocortin-3-receptor (MC-3R) was

```


■■■■■

PI Morsey MA, Sheppard

PI Morsey MA, Sheppard MG;

XX WPI; 2001-026585/04.
 XX New polypeptide variants of growth hormone releasing hormone with
 PT enhanced resistance to enzymatic degradation, useful for treating
 PT growth hormone deficiency related disorders or to improve growth and
 PT performance -
 XX
 PS Examples; Page 41-43; 67pp; English.
 XX The present invention relates to growth hormone-releasing hormone
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.
 CC The variant GHRH polypeptides can be administered to animals to
 CC treat growth hormone deficiency related disorders, or to improve
 CC growth and/or performance. The variants can be included in
 CC pharmaceutical compositions to promote expression and elevation of
 CC growth hormone. The variants can be produced recombinantly at much
 CC higher levels than prior art variants modified using traditional
 CC chemical methods. They have enhanced resistance to enzymatic
 CC degradation, therefore have increased length of activity.
 XX
 SQ Sequence 5108 BP; 1279 A; 1313 C; 1230 G; 1286 T; 0 other;
 Query Match 6.7%; Score 171; DB 22; Length 5108;
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactctgttcttccatgggtctttct 120
 DB 4723 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactctgttcttccatgggtctttct 4782
 QY 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120
 DB 4783 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 4842
 QY 121 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 171
 DB 4843 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 4893
 RESULT 12
 AAC86262
 ID AAC86262 standard; DNA; 5111 BP.
 AC AAC86262;
 XX
 XX CMV IE promoter enhancer elements.
 DE
 XX GHRH; growth hormone-releasing hormone; enzyme degradation; ds.
 KW
 XX Synthetic.
 OS
 XX EP1052286-A2.
 PN
 XX 15-NOV-2000.
 PD
 XX 12-APR-2000; 2000EP-0302790.
 PF
 XX 12-APR-1999; 99US-0128830.
 PR
 XX (PFIZ) PFIZER PROD INC.
 PA
 XX Morsey MA, Sheppard MG;
 PI
 XX WPI; 2001-026585/04.
 DR
 XX New polypeptide variants of growth hormone releasing hormone with
 PT enhanced resistance to enzymatic degradation, useful for treating
 PT growth hormone deficiency related disorders or to improve growth and
 PT performance -
 PT

XX Examples; Page 47-49; 67pp; English.
 XX The present invention relates to growth hormone-releasing hormone
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.
 CC The variant GHRH polypeptides can be administered to animals to
 CC treat growth hormone deficiency related disorders, or to improve
 CC growth and/or performance. The variants can be included in
 CC pharmaceutical compositions to promote expression and elevation of
 CC growth hormone. The variants can be produced recombinantly at much
 CC higher levels than prior art variants modified using traditional
 CC chemical methods. They have enhanced resistance to enzymatic
 CC degradation, therefore have increased length of activity.
 XX
 SQ Sequence 5111 BP; 1280 A; 1314 C; 1230 G; 1287 T; 0 other;
 Query Match 6.7%; Score 171; DB 22; Length 5111;
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactctgttcttccatgggtctttct 120
 DB 4723 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactctgttcttccatgggtctttct 4782
 QY 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120
 DB 4783 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 4842
 QY 121 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 171
 DB 4843 gcagtcacgctgctgacacgctgtgatcagatatcgcgccgctctagacc 4893
 RESULT 13
 AAC86254
 ID AAC86254 standard; DNA; 5185 BP.
 AC AAC86254;
 XX
 XX 28-FEB-2001 (first entry)
 DT
 XX pGHRH-4 construct.
 DE
 XX GHRH; growth hormone-releasing hormone; enzyme degradation; ds.
 KW
 XX Synthetic.
 OS
 XX EP1052286-A2.
 PN
 XX 15-NOV-2000.
 PD
 XX 12-APR-2000; 2000EP-0302790.
 PF
 XX 12-APR-1999; 99US-0128830.
 PR
 XX (PFIZ) PFIZER PROD INC.
 PA
 XX Morsey MA, Sheppard MG;
 PI
 XX WPI; 2001-026585/04.
 DR
 XX New polypeptide variants of growth hormone releasing hormone with
 PT enhanced resistance to enzymatic degradation, useful for treating
 PT growth hormone deficiency related disorders or to improve growth and
 PT performance -
 XX
 PS Examples; Page 32-34; 67pp; English.
 XX The present invention relates to growth hormone-releasing hormone
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.
 CC The variant GHRH polypeptides can be administered to animals to
 CC treat growth hormone deficiency related disorders, or to improve
 CC

CC growth and/or performance. The variants can be included in
 CC pharmaceutical compositions to promote expression and elevation of
 CC growth hormone. The variants can be produced recombinantly at much
 CC higher levels than prior art variants modified using traditional
 CC chemical methods. They have enhanced resistance to enzymatic
 CC degradation, therefore have increased length of activity.
 XX
 SQ Sequence 5185 BP; 1302 A; 1326 C; 1260 G; 1297 T; 0 other;

Query Match 6.7%; Score 171; DB 22; Length 5185;
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgcggtgctgttaacggtgagggcagctgtagtctgacagactgttcttccatgggtctttct 60
 Db 4723 ttgcggtgctgttaacggtgagggcagctgtagtctgacagactgttcttccatgggtctttct 120
 QY 61 gcgcaccagacataatagctgacagactaaacagactgttcttccatgggtctttct 120
 Db 4783 gcgcaccagacataatagctgacagactaaacagactgttcttccatgggtctttct 4842
 QY 121 gcagtcacgctgctgcacacgtgtgatcagatatcgcgccgctctagacc 171
 Db 4843 gcagtcacgctgctgcacacgtgtgatcagatatcgcgccgctctagacc 4893

RESULT 14
 AAC86266
 ID AAC86266 standard; DNA; 5188 BP.
 XX
 AC AAC86266;
 DT 28-FEB-2001 (first entry)
 DE Plasmid GHRH1-44YTCMV.
 KW GHRH; growth hormone-releasing hormone; enzyme degradation; ds.
 OS Synthetic.
 PN EP1052286-A2.
 PD 15-NOV-2000.
 PF 12-APR-2000; 2000EP-0302790.
 PR 12-APR-1999; 99US-0128830.
 PA (PFIZ) PFIZER PROD INC.
 PI Morsey MA, Sheppard MG;
 XX WPI; 2001-026585/04.

XX New polypeptide variants of growth hormone releasing hormone with
 PT enhanced resistance to enzymatic degradation, useful for treating
 PT growth hormone deficiency related disorders or to improve growth and
 PT performance -
 XX Examples; Page 54-56; 67pp; English.
 XX The present invention relates to growth hormone-releasing hormone
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.
 CC The variant GHRH polypeptides can be administered to animals to
 CC treat growth hormone deficiency related disorders, or to improve
 CC growth and/or performance. The variants can be included in
 CC pharmaceutical compositions to promote expression and elevation of
 CC growth hormone. The variants can be produced recombinantly at much
 CC higher levels than prior art variants modified using traditional
 CC chemical methods. They have enhanced resistance to enzymatic
 CC degradation, therefore have increased length of activity.
 XX

SQ Sequence 5188 BP; 1303 A; 1326 C; 1260 G; 1299 T; 0 other;
 Query Match 6.7%; Score 171; DB 22; Length 5188;
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 ttgcggtgctgttaacggtgagggcagctgtagtctgacagactgttcttccatgggtctttct 60
 Db 4723 ttgcggtgctgttaacggtgagggcagctgtagtctgacagactgttcttccatgggtctttct 120
 QY 61 gcgcaccagacataatagctgacagactaaacagactgttcttccatgggtctttct 120
 Db 4783 gcgcaccagacataatagctgacagactaaacagactgttcttccatgggtctttct 4842
 QY 121 gcagtcacgctgctgcacacgtgtgatcagatatcgcgccgctctagacc 171
 Db 4843 gcagtcacgctgctgcacacgtgtgatcagatatcgcgccgctctagacc 4893

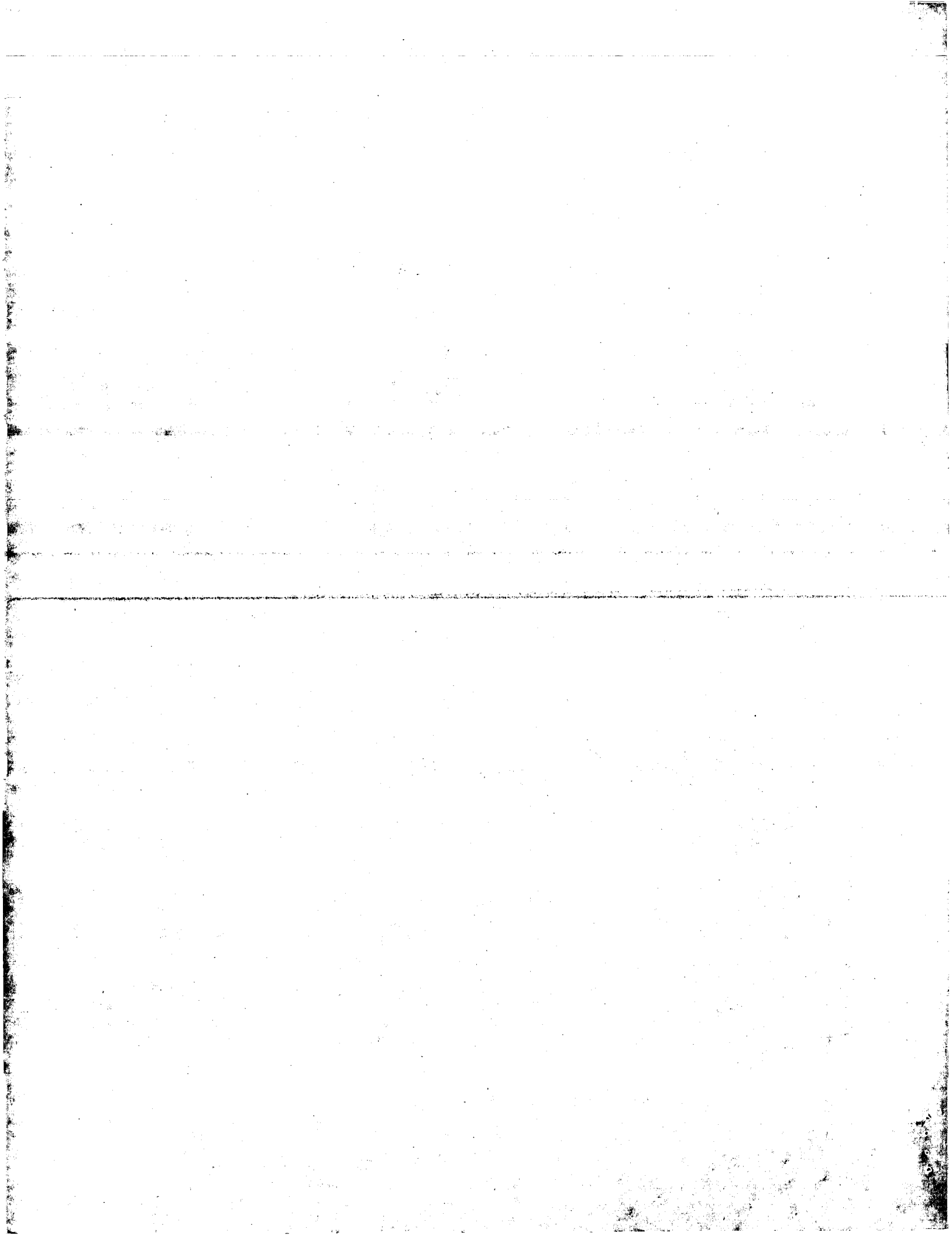
RESULT 15
 AAC86267
 ID AAC86267 standard; DNA; 5254 BP.
 XX
 AC AAC86267;
 DT 28-FEB-2001 (first entry)
 DE Plasmid pGHRH1-44WTCGHPep.
 KW GHRH; growth hormone-releasing hormone; enzyme degradation; ds.
 OS Synthetic.
 PN EP1052286-A2.
 PD 15-NOV-2000.
 PF 12-APR-2000; 2000EP-0302790.
 PR 12-APR-1999; 99US-0128830.
 PA (PFIZ) PFIZER PROD INC.
 PI Morsey MA, Sheppard MG;
 XX WPI; 2001-026585/04.

XX New polypeptide variants of growth hormone releasing hormone with
 PT enhanced resistance to enzymatic degradation, useful for treating
 PT growth hormone deficiency related disorders or to improve growth and
 PT performance -
 XX Examples; Page 56-58; 67pp; English.
 XX The present invention relates to growth hormone-releasing hormone
 CC (GHRH) variants having enhanced resistance to enzymatic degradation.
 CC The variant GHRH polypeptides can be administered to animals to
 CC treat growth hormone deficiency related disorders, or to improve
 CC growth and/or performance. The variants can be included in
 CC pharmaceutical compositions to promote expression and elevation of
 CC growth hormone. The variants can be produced recombinantly at much
 CC higher levels than prior art variants modified using traditional
 CC chemical methods. They have enhanced resistance to enzymatic
 CC degradation, therefore have increased length of activity.
 XX

SQ Sequence 5254 BP; 1318 A; 1334 C; 1285 G; 1317 T; 0 other;
 Query Match 6.7%; Score 171; DB 22; Length 5254;
 Best Local Similarity 100.0%; Pred. No. 1.8e-30;
 Matches 171; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtggaggcagtgtagtctgagcagtagtctgctgcgcgc 60
 Db 4723 ttgcggtgctgttaacggtggaggcagtgtagtctgagcagtagtctgctgcgcgc 4782
 QY 61 gcgccaccagacataatagctgacagactaacagactgttcccttccatgggtctttct 120
 Db 4783 gcgccaccagacataatagctgacagactaacagactgttcccttccatgggtctttct 4842
 QY 121 gcagtcaccgtctgacacgctgtgtgatcagatatcgcgccgctctagacc 171
 Db 4843 gcagtcaccgtctgacacgctgtgtgatcagatatcgcgccgctctagacc 4893

Search completed: February 7, 2002, 19:46:11
 Job time: 14285 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 243.49 seconds
(without alignments)
2367.188 Million cell updates/sec

Title: US-09-391-606-1

Perfect score: 2545

Sequence: 1 ttgcgggtgctttaacgggtg.....tgccactccactgtccttt 2545

Scoring table: IDENTITY_NUC

Gapop 10.0, Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	149.2	5.9	4928	1	US-08-345-913-1
2	149.2	5.9	4928	3	US-08-818-562-1
3	146.8	5.8	150	4	US-08-659-473-8
4	142	5.6	3125	2	US-08-037-816A-13
5	142	5.6	3125	2	US-08-530-146-13
6	139.8	5.5	4328	4	US-09-132-808-1
7	139.8	5.5	4328	4	US-08-910-647-2
8	139.8	5.5	4818	4	US-08-910-647-4
9	139.8	5.5	5107	4	US-08-910-647-3
10	139.8	5.5	9600	4	US-08-910-647-1
11	138.2	5.4	5676	2	US-08-663-998-3
12	138.2	5.4	5682	2	US-08-663-998-4
13	138.2	5.4	5900	2	US-08-663-998-1
14	138.2	5.4	5952	2	US-08-663-998-2
15	138	5.4	2057	1	US-08-450-945-57
16	138	5.4	2057	1	US-08-976-161-57
17	137.8	5.4	13254	1	US-08-276-852-156
18	137.8	5.4	13254	1	US-08-276-852-170
19	137.8	5.4	13254	1	US-08-899-575-156
20	137.8	5.4	13254	1	US-08-899-575-170
21	137.8	5.4	13254	1	US-08-899-575-156
22	137.8	5.4	13254	1	US-08-899-575-170
23	137.8	5.4	13254	5	PCT-US95-08743-156
24	137.8	5.4	13254	5	PCT-US95-08743-170
25	132	5.2	4326	4	US-08-760-615-7
26	111.6	4.4	420	4	US-09-158-863C-64
27	109	4.3	456	1	US-08-450-945-70

Sequence 70, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 19, Appl
Sequence 33, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 1, Appl
Sequence 3, Appl
Sequence 5, Appl
Sequence 5, Appl
Sequence 15, Appl
Sequence 17, Appl
Sequence 19, Appl
Sequence 1, Appl

456 4 US-08-976-161-70
6253 2 US-08-627-151A-5
3853 3 US-08-801-092-5
4026 3 US-08-801-092-19
4249 3 US-08-801-092-33
4283 1 US-08-343-401A-3
4283 1 US-08-445-265A-1
4283 3 US-08-990-442-1
4965 2 US-08-564-313-1
4965 5 PCT-US94-06069-1
5653 1 US-08-073-836-3
5653 1 US-08-235-277-1
6238 3 US-08-646-538-5
6238 4 US-09-503-222-5
6253 3 US-08-893-327-15
6280 3 US-08-893-327-17
6280 3 US-08-893-327-19
6367 1 US-08-470-299-1

ALIGNMENTS

RESULT 1
US-08-345-913-1
; Sequence 1, Application US/08345913
; Patent No. 5641665
; GENERAL INFORMATION:
; APPLICANT: Hobart, Peter
; APPLICANT: Parker, Suzanne
; APPLICANT: Margalith, Michal
; APPLICANT: Khatibi, Shirin
; TITLE OF INVENTION: PLASMIDS SUITABLE FOR IL-2 EXPRESSION
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Knobber, Martens, Olson and Bear
; STREET: 620 Newport Center Drive 16th Floor
; CITY: Newport Beach
; STATE: CA
; COUNTRY: USA
; ZIP: 92660
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq Version 1.5
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/345,913
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Ways Vensko, Nancy
; REGISTRATION NUMBER: 36,298
; REFERENCE/DOCKET NUMBER: VICAL.043A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-235-8550
; TELEFAX: 619-235-0176
; TELEX:
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4928 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE:
; ORIGINAL SOURCE:
; FEATURE:

NAME/KEY: Coding Sequence
LOCATION: 1689...2159
OTHER INFORMATION:
US-08-345-913-1

Query Match 5.9%; Score 149.2; DB 1; Length 4928;
Best Local Similarity 95.1%; Pred. No. 1.7e-30;
Matches 154; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 ggtgtgttaacggtgagggcagtgtagtctgagcagtagctgtgtcgcgcgcgcgc 64
Db 1500 gttgctgttaacggtgagggcagtgtagtctgagcagtagctgtgtcgcgcgcgcgc 1559

QY 65 caccagacataatagctgacagactaacagactgttcttccatgggtcttttttcgcag 124
Db 1560 caccagacataatagctgacagactaacagactgttcttccatgggtcttttttcgcag 1619

QY 125 tcaccgtgacagctgtgatcagatcgcggccgctct 166
Db 1620 TCACCgtgacagctgtgatcagatcgcggccgctctCAACCCT 1661

RESULT 2
US-08-818-562-1
Sequence 1, Application US/08818562
Patent No. 6147055
GENERAL INFORMATION:
APPLICANT: Hobart, Peter M.
APPLICANT: Margalith, Michal
APPLICANT: Parker, Suzanne E.
APPLICANT: Khatibi, Shirin
TITLE OF INVENTION: Plasmids Suitable for IL-2 Expression
FILE REFERENCE: 1530.0080001
CURRENT APPLICATION NUMBER: US/08/818,562
CURRENT FILING DATE: 1997-03-14
EARLIER APPLICATION NUMBER: US 08/345,913
EARLIER FILING DATE: 1994-11-28
NUMBER OF SEQ ID NOS: 3
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 1
LENGTH: 4928
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1689)...(2159)
US-08-818-562-1

Query Match 5.9%; Score 149.2; DB 3; Length 4928;
Best Local Similarity 95.1%; Pred. No. 1.7e-30;
Matches 154; Conservative 0; Mismatches 8; Indels 0; Gaps 0;

QY 5 ggtgtgttaacggtgagggcagtgtagtctgagcagtagctgtgtcgcgcgcgcgc 64
Db 1500 gttgctgttaacggtgagggcagtgtagtctgagcagtagctgtgtcgcgcgcgcgc 1559

QY 65 caccagacataatagctgacagactaacagactgttcttccatgggtcttttttcgcag 124
Db 1560 caccagacataatagctgacagactaacagactgttcttccatgggtcttttttcgcag 1619

QY 125 tcaccgtgacagctgtgatcagatcgcggccgctct 166
Db 1620 TCACCgtgacagctgtgatcagatcgcggccgctctCAACCCT 1661

RESULT 3
US-08-659-473-8
Sequence 8, Application US/08659473
Patent No. 6210876
GENERAL INFORMATION:
APPLICANT: M. B. Cerney

TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: Abbott Laboratories
STREET: 100 Abbott Park Road
CITY: Abbott Park
STATE: Illinois
COUNTRY: USA
ZIP: 60064-3500
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: System 7.0.1
SOFTWARE: Microsoft Word 5.1a
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/659,473
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Paul D. Yasger
REGISTRATION NUMBER: 37,477
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE: 708/937-2341
TELEFAX: 708/938-2623
TELEX:
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 150 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA (C. pneumoniae)
US-08-659-473-8

Query Match 5.8%; Score 146.8; DB 4; Length 150;
Best Local Similarity 98.7%; Pred. No. 1.2e-30;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 1305 tacctcaacatcactagctgacatcacaggtgctttgtgagcctccaggatgcttcac 1364
Db 1 TACCTCAACATCAGTCTGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCTGTAC 60

QY 1365 taataaagatcacagcggctactgtaggaaaccccaatcgctcgaggtgggaaac 1424
Db 61 TAATATAAGGATACAGCGGCTACTGTGATGAGGAAACCCCAATCGCTGCGGTGTGGGAAAC 120

QY 1425 taagaatccgatgcaattaaagtggcgc 1454
Db 121 TAAGAATCCGATGATGAGTTAAAGTTGGCGC 150

RESULT 4
US-08-037-816A-13
Sequence 13, Application US/08037816A
Patent No. 5869624
GENERAL INFORMATION:
APPLICANT: Hasel, Karl W.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
THERETO, AND THERAPEUTIC AND PROPHYLACTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM: Floppy disk
MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/037,816A
FILING DATE: 26-MAR-1993
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
TELEPHONE: (212) 977-9550
TELEFAX: (212) 977-9550
INFORMATION FOR SEQ ID NO: 13:
LENGTH: 3125 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1555..3115
OTHER INFORMATION:
US-08-037-816A-13

Query Match 5.6%; Score 142; DB 2; Length 3125;
Best Local Similarity 91.0%; Pred. No. 1.2e-28;
Matches 151; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactcgttgcgcgcgc 60
|||||
Db 1414 TTGCGGTGCTGTAAACGGTGGAGGCAGTGTAGTCTGAGCAGTACTCGTTGTCGCGCGC 1473

QY 61 ggcgcacacagacataatagctgacagactaacagactgttccttccatgggtctttct 120
|||||
Db 1474 GCGCCACCACAGACATAATAGCTGACAGACTTAACAGACTGTTCCTTCCATGGGTCTTTCT 1533

QY 121 cgagtcaccgtgctgtgacagctgtgatcagatcgcgcgcgcctct 166
|||||
Db 1534 GCAGTCACCGTCTCTTGACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1579

RESULT 5
US-08-146-13
Sequence 13, Application US/08530146
Patent No. 5886163
GENERAL INFORMATION:
APPLICANT: Hasel, Karl W.
APPLICANT: Maddon, Paul J.
TITLE OF INVENTION: HIV-1 VACCINES, ANTIBODY COMPOSITIONS RELATED
TITLE OF INVENTION: THEREOF, AND THERAPEUTIC AND PROPHYLACTIC USES
TITLE OF INVENTION: THEREOF
NUMBER OF SEQUENCES: 29
CORRESPONDENCE ADDRESS:
ADDRESSEE: Cooper & Dunham
STREET: 30 Rockefeller Plaza
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10112
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.24
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/530,146
FILING DATE:
CLASSIFICATION:

PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/037,816
FILING DATE: 26-MAR-1993
ATTORNEY/AGENT INFORMATION:
NAME: White, John P.
REGISTRATION NUMBER: 28,678
REFERENCE/DOCKET NUMBER: 41190/JPW/AJM
TELEPHONE: (212) 977-9550
TELEFAX: (212) 664-0525
TELEX: 422523 COOPUI
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 3125 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1555..3115
OTHER INFORMATION:
US-08-530-146-13

Query Match 5.6%; Score 142; DB 2; Length 3125;
Best Local Similarity 91.0%; Pred. No. 1.2e-28;
Matches 151; Conservative 0; Mismatches 15; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtctgagcagactcgttgcgcgcgc 60
|||||
Db 1414 TTGCGGTGCTGTAAACGGTGGAGGCAGTGTAGTCTGAGCAGTACTCGTTGTCGCGCGC 1473

QY 61 ggcgcacacagacataatagctgacagactaacagactgttccttccatgggtctttct 120
|||||
Db 1474 GCGCCACCACAGACATAATAGCTGACAGACTTAACAGACTGTTCCTTCCATGGGTCTTTCT 1533

QY 121 cgagtcaccgtgctgtgacagctgtgatcagatcgcgcgcgcctct 166
|||||
Db 1534 GCAGTCACCGTCTCTTGACACGATGATGATGATGATGATGATGATGATGATGATGATGAT 1579

RESULT 6
US-09-132-803-1
Sequence 1, Application US/09132808
Patent No. 6197332
GENERAL INFORMATION:
APPLICANT: Ronald Zuckermann et al.
TITLE OF INVENTION: Lipid-Conjugated Polyamide Compounds and Related
TITLE OF INVENTION: Compositions and Methods Thereof
NUMBER OF SEQUENCES: 1
CORRESPONDENCE ADDRESS:
ADDRESSEE: Chiron Corporation
STREET: 4560 Horton Street
CITY: Emeryville
STATE: California
COUNTRY: U.S.A.
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/132,808
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Fujita, Sharon M.
REGISTRATION NUMBER: 38,459
REFERENCE/DOCKET NUMBER: 1387,002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2706

TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-09-132-808-1

Query Match 5.5%; Score 139.8; DB 4; Length 4328;
Best Local Similarity 95.4%; Pred. No. 5.6e-28;
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtgagcagactgttccttcctggtctttct 60
Db 1511 TTGCGGTGCTGTTAAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGCGCGCGC 1570
QY 61 gcgccaccagacataatagctgacagactaacagactgttccttcctggtctttct 120
Db 1571 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTCCATGGGTCTTTCT 1630
QY 121 gcagtcaccgtctgcagacgctgtgatcaga 151
Db 1631 GCAGTCACCGCTGCTCGACCTAAGATTTCAGA 1661

RESULT 7
US-08-910-647-2
; Sequence 2, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4328 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-2

Query Match 5.5%; Score 139.8; DB 4; Length 4328;
Best Local Similarity 95.4%; Pred. No. 5.6e-28;
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagtgagcagactgttccttcctggtctttct 60
Db 1511 TTGCGGTGCTGTTAAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGCGCGCGC 1570
QY 61 gcgccaccagacataatagctgacagactaacagactgttccttcctggtctttct 120
Db 1571 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTCCATGGGTCTTTCT 1630
QY 121 gcagtcaccgtctgcagacgctgtgatcaga 151
Db 1631 GCAGTCACCGCTGCTCGACCTAAGATTTCAGA 1661
RESULT 8
US-08-910-647-4
; Sequence 4, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910,647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 4818 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-4

Query Match 5.5%; Score 139.8; DB 4; Length 4818;
Best Local Similarity 95.4%; Pred. No. 6e-28;
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;
QY 1 ttgcggtgctgttaacggtgagggcagtgtagtgagcagactgttccttcctggtctttct 60
Db 1514 TTGCGGTGCTGTTAAACGGTGGAGGCGAGTGTAGTCTGAGCAGTACTCGTTGCGCGCGC 1573
QY 61 gcgccaccagacataatagctgacagactaacagactgttccttcctggtctttct 120
Db 1574 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTCCATGGGTCTTTCT 1633
QY 121 gcagtcaccgtctgcagacgctgtgatcaga 151
Db 1634 GCAGTCACCGCTGCTCGACCTAAGATTTCAGA 1664

RESULT 9

US-08-910-647-3
; Sequence 3, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California
; COUNTRY: U.S.A.
; ZIP: 94608-2916
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/910.647
; FILING DATE:
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Fujita, Sharon M.
; REGISTRATION NUMBER: 38,459
; REFERENCE/DOCKET NUMBER: 1218.002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (510) 923-2706
; TELEFAX: (510) 655-3542
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5107 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-910-647-3

Query Match 5.5%; Score 139.8; DB 4; Length 5107;
Best Local Similarity 95.4%; Pred. No. 6.1e-28;
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ttgcggtctgttaacgtgagggcagctgtagctgacagactgttcttccatgggtctttct 60
|||||
Db 1514 TTGCGGTGCTGTTAACGGTGGAGGCGAGTGTAGCTGACAGACTGTTCTTTCCATGGGCTTTTCT 1573
Qy 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120
|||||
Db 1574 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTTCCATGGGCTTTTCT 1633
Qy 121 gcagtcacgtgctgacacgtgtgatcaga 151
|||||
Db 1634 GCAGTCACCGTGGTGCAGCTAAGAAATTCAGA 1664

RESULT 10

US-08-910-647-1
; Sequence 1, Application US/08910647
; Patent No. 6251433
; GENERAL INFORMATION:
; APPLICANT: Zuckermann et al.
; TITLE OF INVENTION: Compositions and Methods for
; TITLE OF INVENTION: Polynucleotide Delivery
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Chiron Corporation
; STREET: 4560 Horton Street
; CITY: Emeryville
; STATE: California

COUNTRY: U.S.A.
ZIP: 94608-2916
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910.647
FILING DATE:
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Fujita, Sharon M.
REGISTRATION NUMBER: 38,459
REFERENCE/DOCKET NUMBER: 1218.002
TELECOMMUNICATION INFORMATION:
TELEPHONE: (510) 923-2706
TELEFAX: (510) 655-3542
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9600 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-910-647-1

Query Match 5.5%; Score 139.8; DB 4; Length 9600;
Best Local Similarity 95.4%; Pred. No. 8.6e-28;
Matches 144; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 ttgcggtctgttaacgtgagggcagctgtagctgacagactgttcttccatgggtctttct 60
|||||
Db 6616 TTGCGGTGCTGTTAACGGTGGAGGCGAGTGTAGCTGACAGACTGTTCTTTCCATGGGCTTTTCT 6675
Qy 61 gcgccaccagacataatagctgacagactaacagactgttcttccatgggtctttct 120
|||||
Db 6676 GCGCCACCAGACATAATAGCTGACAGACTAACAGACTGTTCTTTCCATGGGCTTTTCT 6735
Qy 121 gcagtcacgtgctgacacgtgtgatcaga 151
|||||
Db 6736 GCAGTCACCGTGGTGCAGCTAAGAAATTCAGA 6766

RESULT 11

US-08-663-998-3
; Sequence 3, Application US/08663998
; Patent No. 5846946
; GENERAL INFORMATION:
; APPLICANT: HUEBNER, Robert C.
; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent in Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/663.998

```

; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5676 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-663-998-3

Query Match          5.4%; Score 138.2; DB 2; Length 5676;
Best Local Similarity 97.9%; Pred. No. 1.8e-27;
Matches 140; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagctgagcagtagctgttccatggtgtttct 120
Db 1549 TTGCGGTGCTGTAAACGCTGGAGGCGAGTGTAGCTGAGCAGTACTCTGTTGCTGCCGCGC 1608

QY 61 ggcaccacagacataatagctgagcagactaacagactgttccatggtgtttct 120
Db 1609 GCGCCACACAGACATAATAGCTGAGCAGTACTCTGAGCAGTACTCTTCCATGGGTCTTTCT 1668

QY 121 gcagtcaccgtctgcacacgtg 143
Db 1669 GCAGTCACCGTCTGCACACAG 1691

RESULT 12
US-08-663-998-4
; Sequence 4, Application US/08663998
; Patent No. 5846946
; GENERAL INFORMATION:
; APPLICANT: HUEBNER, Robert C.
; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5676 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-663-998-3

Query Match          5.4%; Score 138.2; DB 2; Length 5676;
Best Local Similarity 97.9%; Pred. No. 1.8e-27;
Matches 140; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcggtgctgttaacggtgagggcagtgtagctgagcagtagctgttccatggtgtttct 120
Db 1549 TTGCGGTGCTGTAAACGCTGGAGGCGAGTGTAGCTGAGCAGTACTCTGTTGCTGCCGCGC 1608

QY 61 ggcaccacagacataatagctgagcagactaacagactgttccatggtgtttct 120
Db 1609 GCGCCACACAGACATAATAGCTGAGCAGTACTCTGAGCAGTACTCTTCCATGGGTCTTTCT 1668

QY 121 gcagtcaccgtctgcacacgtg 143
Db 1669 GCAGTCACCGTCTGCACACAG 1691

RESULT 13
US-08-663-998-1
; Sequence 1, Application US/08663998
; Patent No. 5846946
; GENERAL INFORMATION:
; APPLICANT: HUEBNER, Robert C.
; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5900 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-663-998-1

Query Match          5.4%; Score 138.2; DB 2; Length 5900;
Best Local Similarity 97.9%; Pred. No. 1.8e-27;

```

Matches 140; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcggtcgtttaacggtggaggcagtgtagctgagcagtagctcgttgcgcgcgc 60
|||||
Db 1553 TTGCGGTGCTGTTAAACGGTGGAGGCAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGC 1612
QY 61 ggcaccacagacataatagctgagcagactaacagagtagcttccttccatgggtctttct 120
|||||
Db 1613 GCGCCACCACAGACATAATAGCTGACAGACTAACAGAGACTGTTCTTCCATGGGTCTTTTCT 1672
QY 121 gcagtcaccgtcgtcgacacgtg 143
|||||
Db 1673 GCAGTCACCGTCTGTCGACAGAG 1695

RESULT 14

US-08-663-998-2
; Sequence 2, Application US/08663998
; Patent No. 5846946

GENERAL INFORMATION:

; APPLICANT: HUEBNER, Robert C.
; APPLICANT: NORMAN, Jon A.
; APPLICANT: LIANG, Xiaowu
; APPLICANT: CARNER, Kristin R.
; APPLICANT: BARBOUR, Alan G.
; APPLICANT: LUKE, Catherine J.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: ADMINISTERING BORRELIA DNA
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford, P.C.
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/663,998
; FILING DATE: 06-JUN-1996
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Kowalski, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2440
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-840-3333
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 5952 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)

US-08-663-998-2

Query Match

Best Local Similarity 5.4%; Score 138.2; DB 2; Length 5952;

Matches 140; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 ttgcggtcgtttaacggtggaggcagtgtagctgagcagtagctcgttgcgcgcgc 60
|||||
Db 1552 TTGCGGTGCTGTTAAACGGTGGAGGCAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGC 1611
QY 61 ggcaccacagacataatagctgagcagactaacagagtagcttccttccatgggtctttct 120
|||||
Db 1612 GCGCCACCACAGACATAATAGCTGACAGACTAACAGAGACTGTTCTTCCATGGGTCTTTTCT 1671

QY 121 gcagtcaccgtcgtcgacacgtg 143
|||||
Db 1672 GCAGTCACCGTCTGTCGACAGAG 1694

RESULT 15

US-08-450-945-57/c
; Sequence 57, Application US/08450945
; Patent No. 5783383

GENERAL INFORMATION:

; APPLICANT: Kondo, Kazuhiro
; APPLICANT: Mocarski, Edward S. Jr.
; TITLE OF INVENTION: LATENT TRANSCRIPTS AND PROMOTERS
; TITLE OF INVENTION: OF CYTOMEGALOVIRUS
; NUMBER OF SEQUENCES: 75
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Dehlinger & Associates
; STREET: 350 Cambridge Avenue, Suite 250
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94306

COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/450,945
; FILING DATE: 23-MAY-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Sholtz, Charles K.
; REGISTRATION NUMBER: 38,615
; REFERENCE/DOCKET NUMBER: 8600-0157
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 324-0880
; TELEFAX: (415) 324-0960
; INFORMATION FOR SEQ ID NO: 57:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2057 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: cDNA to mRNA
; HYPOTHETICAL: NO
; ANTI-SENSE: YES
; ORIGINAL SOURCE:
; INDIVIDUAL ISOLATE: region of AD169 strain HCMV (antisense)
; INDIVIDUAL ISOLATE: containing antisense transcript ORFs

US-08-450-945-57

Query Match

Best Local Similarity 5.4%; Score 138; DB 1; Length 2057;

Matches 141; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

QY 1 ttgcggtcgtttaacggtggaggcagtgtagctgagcagtagctcgttgcgcgcgc 60
|||||
Db 1632 TTGCGGTGCTGTTAAACGGTGGAGGCAGTGTAGTCTGAGCAGTACTCGTTGCTGCCGCGC 1573
QY 61 ggcaccacagacataatagctgagcagactaacagagtagcttccttccatgggtctttct 120
|||||
Db 1572 GCGCCACCACAGACATAATAGCTGACAGACTAACAGAGACTGTTCTTCCATGGGTCTTTTCT 1513
QY 121 gcagtcaccgtcgtcgacacgtgta 146
|||||
Db 1512 GCAGTCACCGTCTTTCGACACGATGGA 1487

Search completed: February 7, 2002, 15:53:24
Job time: 318 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:48:06 ; Search time 7419.31 Seconds
(without alignments)
3686.058 Million cell updates/sec

Title: US-09-391-606-1

Perfect score: 2545

Sequence: 1 ttgcgggtgcttaacggtg.....tgccactcccaactgtccttt 2545

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estfun:*

2: em_esthum:*

3: em_estin:*

4: em_estom:*

5: em_estpl:*

6: em_estba:*

7: em_estro:*

8: em_estov:*

9: em_htc:*

10: gb_est1:*

11: gb_est2:*

12: gb_htc:*

13: gb_gss:*

14: em_gss_fun:*

15: em_gss_hum:*

16: em_gss_inv:*

17: em_gss_pln:*

18: em_gss_pro:*

19: em_gss_rtd:*

20: em_gss_vrt:*

21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	87	3.4	149	10	AW669676
2	87	3.4	162	10	AW669522
3	87	3.4	162	10	AW785253
4	87	3.4	162	10	AW785901
5	87	3.4	243	10	AW669518
6	87	3.4	503	11	BF074689
7	85.4	3.4	328	10	AW669634
8	84.4	3.3	374	10	AW669633
9	71	2.8	98	11	BG365021
10	61.2	2.4	1001	13	CNS007BE
11	60	2.4	1101	13	CNS0039G
12	59.4	2.3	996	13	CNS00FUH

C 13	56	2.2	987	13	CNS014PQ
C 14	53.8	2.1	976	13	CNS04E5M
C 15	52.4	2.1	240	10	AU071789
C 16	52.2	2.1	928	13	CNS00DKY
C 17	52.2	2.1	937	13	CNS006XP
C 18	51.4	2.0	663	13	BH001815
C 19	51.2	2.0	1190	13	CNS020N7
C 20	51	2.0	1082	13	CNS012XZ
C 21	51	2.0	1092	13	CNS020K7
C 22	51	2.0	1101	13	CNS00FYG
C 23	50.8	2.0	1001	13	CNS0155H
C 24	50.6	2.0	1225	13	CNS0161D
C 25	50.4	2.0	842	13	AZ670257
C 26	50.4	2.0	1101	13	CNS003B0
C 27	50.2	2.0	822	13	CNS009CW
C 28	50.2	2.0	1101	13	CNS00K85
C 29	50.2	2.0	1225	13	CNS0161D
C 30	50	2.0	842	13	CNS010QN
C 31	49.8	2.0	942	13	CNS018GS
C 32	49.8	2.0	1101	13	CNS000B8
C 33	49.6	1.9	668	10	AL514901
C 34	49.4	1.9	777	13	CNS025WB
C 35	49.4	1.9	1083	13	CNS009YQ
C 36	49.2	1.9	408	13	AQ954772
C 37	49.2	1.9	1131	13	CNS034FO
C 38	49.2	1.9	1203	13	CNS015WU
C 39	49	1.9	905	13	CNS00KHX
C 40	48.8	1.9	1101	13	CNS00LVZ
C 41	48.8	1.9	1101	13	CNS016DT
C 42	48.6	1.9	838	13	CNS03272
C 43	48.6	1.9	988	13	CNS0161M
C 44	48.4	1.9	773	13	AZ183506
C 45	48.4	1.9	945	13	CNS04D0K

ALIGNMENTS

RESULT 1

LOCUS: AW669676 149 bp mRNA EST 25-APR-2001
 DEFINITION: 113173 MARC lBOV Bos taurus cDNA 5', mRNA sequence.
 ACCESSION: AW669676
 VERSION: AW669676.1 GI:7526190
 KEYWORDS: EST.
 SOURCE: COW.
 ORGANISM: Bos taurus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae; Bovidae; Bovinae; Bos.
 Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T., Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett, G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G., Perle, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and Keefe, J.W.

TITLE: Sequence evaluation of four pooled-tissue normalized bovine cDNA libraries and construction of a gene index for cattle
 JOURNAL: Genome Res. 11 (4), 626-630 (2001)
 MEDLINE: 21180013
 COMMENT: Contact: Smith TPL
 USDA, ARS, US Meat Animal Research Center
 PO Box 166, Clay Center, NE 68933-0166, USA
 Tel: 402 762 4366
 Fax: 402 762 4390
 Email: smitht@mail.marc.usda.gov

Single pass sequencing. Bases called and alt_trimmed with phred v0.980904.e. Vector identified by cross_match with the -mnscore 18 and -minmatch 12 options.
 PCR Primers
 FORWARD: AGGAAACAGCTATGACCAT
 BACKWARD: GTTTTCCCGTCAGCAGC
 Plate: 107 row: D column: 22

```
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..149
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC IBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
24 a 48 c 38 g 39 t

BASE COUNT
ORIGIN

Query Match 3.4%; Score 87; DB 10; Length 149;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtacctctagtgcagccatctgtttgttgcctcccccgtgccttcttgacc 2518
|||||
Db 36 CTGTGCTTCTAGTTCACAGCATCTGTGTGCTCCCTCCCGTGCCTTCTTGACCC 95
|||||

QY 2519 tggaaagtgcactccactgtccttt 2545
|||||
Db 96 TGGAGGTGCCACTCCCACTGCTCTT 122
|||||

RESULT 2
AW669522
LOCUS 112928 MARC IBOV Bos taurus cDNA 5', mRNA sequence. 25-APR-2001
DEFINITION AW669522
ACCESSION AW669522
VERSION AW669522.1 GI:7526036
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 162)
Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrrenkrug, S.C., Bennett
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-Mckown, C.G.,
Perle, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keel, J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 106 row: K column: 23
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..162
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC IBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
```

```
fat, hypothalamus, and pituitary."
28 a 51 c 42 g 41 t

BASE COUNT
ORIGIN

Query Match 3.4%; Score 87; DB 10; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtacctctagtgcagccatctgtttgttgcctcccccgtgccttcttgacc 2518
|||||
Db 49 CTGTGCTTCTAGTTCACAGCATCTGTGTGCTCCCTCCCGTGCCTTCTTGACCC 108
|||||

QY 2519 tggaaagtgcactccactgtccttt 2545
|||||
Db 109 TGGAGGTGCCACTCCCACTGCTCTT 135
|||||

RESULT 3
AW785253
LOCUS 116166 MARC 1P1G Sus scrofa cDNA 5', mRNA sequence. 09-JUL-2000
DEFINITION AW785253
ACCESSION AW785253
VERSION AW785253.1 GI:7842029
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 162)
Fahrrenkrug, S.C., Freking, B.A., Rohrer, G.A., Smith, T.P.L., Casas, E.,
Stone, R.T., Heaton, M.P., Grosse, W.M., Bennett, G.A., Laegreid, W.W.,
and Keel, J.W.
Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
Unpublished (2000)
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCACGTCACGACG
Plate: 34 row: K column: 23
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers
1..162
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 1P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
28 a 51 c 42 g 41 t

BASE COUNT
ORIGIN

Query Match 3.4%; Score 87; DB 10; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtacctctagtgcagccatctgtttgttgcctcccccgtgccttcttgacc 2518
|||||
Db 49 CTGTGCTTCTAGTTCACAGCATCTGTGTGCTCCCTCCCGTGCCTTCTTGACCC 108
|||||

QY 2519 tggaaagtgcactccactgtccttt 2545
```

```

|||||
Db 109 TGGAAAGTGCCACTCCCACTGTCCTTT 135

RESULT 4
LOCUS AW785901 162 bp mRNA EST 09-JUL-2000
DEFINITION 117399 MARC lPIG Sus scrofa cDNA 5', mRNA sequence.
ACCESSION AW785901
VERSION AW785901.1 GI:7842677
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
REFERENCE 1 (bases 1 to 162)
AUTHORS Fahrenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Laegreid,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 37 row: K column: 23
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1..162
/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC lPIG"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from day 11, 13, 15, 20,
and 30 embryos."
BASE COUNT 28 a 51 c 42 g 41 t
ORIGIN
Query Match 3.4%; Score 87; DB 10; Length 162;
Best Local Similarity 100.0%; Pred. No. 1.3e-07;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtgccttctagtgcagccatctgtttgttgcctcccccgtgcttcccttgacc 2518
|||||
Db 49 CTGTGCCCTCTAGTTCGCCAGCATCTGTTGTGCCCTCCCGTCCCTCTCTTGACCC 108

QY 2519 tggaagggtgccactccactgtccttt 2545
|||||
Db 109 TGGAAAGTGCCACTCCCACTGTCCTTT 135

RESULT 5
LOCUS AW669518 243 bp mRNA EST 25-APR-2001
DEFINITION 112921 MARC lBOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION AW669518
VERSION AW669518.1 GI:7526032
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 503)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 243)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smithemail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGACG
Plate: 106 row: J column: 22
Seq primer: ATTTAGGTGACACTATAG.
FEATURES Location/Qualifiers
source 1..243
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC lBOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 56 a 67 c 59 g 61 t
ORIGIN
Query Match 3.4%; Score 87; DB 10; Length 243;
Best Local Similarity 100.0%; Pred. No. 1.2e-07;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtgccttctagtgcagccatctgtttgttgcctcccccgtgcttcccttgacc 2518
|||||
Db 49 CTGTGCCCTCTAGTTCGCCAGCATCTGTTGTGCCCTCCCGTCCCTCTCTTGACCC 108

QY 2519 tggaagggtgccactccactgtccttt 2545
|||||
Db 109 TGGAAAGTGCCACTCCCACTGTCCTTT 135

RESULT 6
LOCUS BF074689 503 bp mRNA EST 25-APR-2001
DEFINITION 222166 MARC 2BOV Bos taurus cDNA 5', mRNA sequence.
ACCESSION BF074689
VERSION BF074689.1 GI:10868200
KEYWORDS EST.
SOURCE cow.
ORGANISM Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
REFERENCE 1 (bases 1 to 503)
AUTHORS Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W.W., Rohrer,G.A., Chitko-Mckown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keele,J.W.
TITLE Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

```

JOURNAL
MEDLINE
COMMENT

Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 82 row: E column: 4
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers

FEATURES
source

1..503
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 2BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, thymus,
semitendinosus muscle, longissimus muscle, pancreas,
adrenal, and endometrium."
BASE COUNT 122 a 146 c 130 g 105 t
ORIGIN

Query Match 3.4%; Score 87; DB 11; Length 503;
Best Local Similarity 100.0%; Pred. No. 1.1e-07;
Matches 87; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2459 ctgtcccttctagttgcagccatctgtttgtccctcccccgtccttctcttgacc 2518
|||||
Db 367 CTGTGCTCTCTAGTTGCCAGCCATCTGTTGTGCCCTCCCGCTCCCTTGTGACCC 426
|||||
QY 2519 tggaggtgcccactcccactgtccttt 2545
|||||
Db 427 TGGAGGTGCCACTCCCACTGTCCTT 453
|||||

RESULT 7
AW669634/c

LOCUS
DEFINITION 113105 MARC 1BOV Bos taurus CDNA 5', mRNA sequence. EST 25-APR-2001
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

REFERENCE
AUTHORS

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

JOURNAL
MEDLINE
COMMENT

Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

v0.980904.e. Vector identified by cross_match with the -minscore 18
and -mismatch 12 options.

PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCCTCAGTCACGAGC
Plate: 107 row: I column: 14
Seq primer: ATTTAGGTGACACTATAG.
Location/Qualifiers

FEATURES
source

1..328
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: PCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
BASE COUNT 71 a 101 g 68 t
ORIGIN

Query Match 3.4%; Score 85.4; DB 10; Length 328;
Best Local Similarity 98.9%; Pred. No. 2.4e-07;
Matches 86; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2459 ctgtcccttctagttgcagccatctgtttgtccctcccccgtccttctcttgacc 2518
|||||
Db 110 CTGTGCTCTCTAGTTGCCAGCCATCTGTTGTGCCCTCCCGCTCCCTTGTGACCC 51
|||||

QY 2519 tggaggtgcccactcccactgtccttt 2545
|||||

Db 50 TGGAGGTGCCACTCCCACTGTCCTT 24
|||||

RESULT 8

AW669633/c
LOCUS
DEFINITION 113104 MARC 1BOV Bos taurus CDNA 5', mRNA sequence. EST 25-APR-2001
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM

REFERENCE

USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

REFERENCE
AUTHORS

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

JOURNAL

MEDLINE

COMMENT

Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

TITLE

Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle

JOURNAL

MEDLINE

COMMENT

Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@email.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred

REFERENCE

AUTHORS

Smith, T.P.L., Grosse, W.M., Freking, B.A., Roberts, A.J., Stone, R.T.,
Casas, E., Wray, J.E., White, J., Cho, J., Fahrenkrug, S.C., Bennett,
G.L., Heaton, M.P., Laegreid, W.W., Rohrer, G.A., Chitko-McKown, C.G.,
Perteau, G., Holt, I., Karamycheva, S., Liang, F., Quackenbush, J. and
Keele, J.W.

TITLE

JOURNAL

MEDLINE

COMMENT


```

/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
79 t
BASE COUNT      84 a 102 c 109 g
ORIGIN

Query Match      3.3%; Score 84.4; DB 10; Length 374;
Best Local Similarity 98.8%; Pred. No. 3.8e-07;
Matches 85; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 2459 cttgctcttagtcagccatctgtttgttgcctccctccctgcttcttgcaccc 2518
Db 109 CTGTGCTTCTAGTCCAGCATCTGTGTTGCCCCCTCCCGGCTTCTCTTGACCC 50

QY 2519 tgaagtgccactccactgctctt 2544
Db 49 TGAAGGTGCCACTCCCACTGTCTT 24

RESULT 9
BG365021      98 bp mRNA EST 25-APR-2001
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Bos taurus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
Bovidae; Bovinae; Bos.
1 (bases 1 to 98)
Smith,T.P.L., Grosse,W.M., Freking,B.A., Roberts,A.J., Stone,R.T.,
Casas,E., Wray,J.E., White,J., Cho,J., Fahrenkrug,S.C., Bennett
,G.L., Heaton,M.P., Laegreid,W., Rohrer,G.A., Chitko-McKown,C.G.,
Perteau,G., Holt,I., Karamycheva,S., Liang,F., Quackenbush,J. and
Keefe,J.W.
Sequence evaluation of four pooled-tissue normalized bovine cDNA
libraries and construction of a gene index for cattle
Genome Res. 11 (4), 626-630 (2001)
21180013
Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt_trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR Primers
FORWARD: AGGAACAGCTATGACCAT
BACKWARD: GTTTCAGTCACGACG
Plate: 72 row: L column: 16
Seq primer: ATTTAGTGACACTATAG.
Location/Qualifiers
1..98
/organism="Bos taurus"
/db_xref="taxon:9913"
/clone_lib="MARC 1BOV"
/tissue_type="pooled"
/lab_host="DH10B"
/Note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from lymph node, ovary,
fat, hypothalamus, and pituitary."
18 a 35 c 18 g 27 t
BASE COUNT
ORIGIN

Query Match      2.4%; Score 61.2; DB 13; Length 1001;
Best Local Similarity 25.4%; Pred. No. 0.017;
Matches 87; Conservative 106; Mismatches 149; Indels 0; Gaps 0;

QY 599 gggtaatttataaagtttgggtgcgcgaagcaattgctctaactatgacctaaag 658
Db 391 GAAGRTTAAARKKTKTKTKTSATTKKCAARSYTTYSRAAWKSTWYMTWTWMMGG 332

QY 659 aggctaattttcaaaattataaaggattttccgctctatatgtatggtcacacag 718
Db 331 GGSWTKGGGCTKGGWKGAACCCGRARGAWTTTCCCTCCWWRCCMWKGCCTKGGRRCS 272

QY 719 attctatcattcagagcttatggggttactccaaaccagcaaacgataaactggca 778

```

```

Query Match      2.8%; Score 71; DB 11; Length 98;
Best Local Similarity 100.0%; Pred. No. 0.00025;
Matches 71; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2475 ccagccatctgtttgcccctcccccgtgcttcttgcacctgagaggtgacctcc 2534
Db 1 CCAGCATCTGTGTTGCCCCCTCCCGCTTCTTGACCTGGAAGGTGCCACTCC 60

QY 2535 cactgtccttt 2545
Db 61 CACTGTCTCTT 71

```

```

RESULT 10
CNS007BE/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
fruit fly.
Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1001)
Direct Submission
Submitted (02-JUN-1999) Genoscope - Centre National de Sequenage :
Rp 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
Location/Qualifiers
1..1001
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR15H24"
/Note="end : TET3"
BASE COUNT      159 a 148 c 151 g 212 t 331 others
ORIGIN

```

```

FEATURES
source

```

```

BASE COUNT      159 a 148 c 151 g 212 t 331 others
ORIGIN

```

```

Query Match      2.4%; Score 61.2; DB 13; Length 1001;
Best Local Similarity 25.4%; Pred. No. 0.017;
Matches 87; Conservative 106; Mismatches 149; Indels 0; Gaps 0;

QY 599 gggtaatttataaagtttgggtgcgcgaagcaattgctctaactatgacctaaag 658
Db 391 GAAGRTTAAARKKTKTKTKTSATTKKCAARSYTTYSRAAWKSTWYMTWTWMMGG 332

QY 659 aggctaattttcaaaattataaaggattttccgctctatatgtatggtcacacag 718
Db 331 GGSWTKGGGCTKGGWKGAACCCGRARGAWTTTCCCTCCWWRCCMWKGCCTKGGRRCS 272

QY 719 attctatcattcagagcttatggggttactccaaaccagcaaacgataaactggca 778

```

```

Db 271 GRWTTTATTTTWWTTTWWTTTACCTTAANKGSGGKTWRWKRKGGKKAATWTRCCOC 212
Qy 779 gtgatttacttcgaaagttgctaggtatataatttcacgctgttaagtcacatttt 838
Db 211 MAAAKTTWASWTTTGGKTTTAAWTTWAKGTWAWKTKTKGTYCYWAAWYKTTTK 152
Qy 839 aataaaatcttaaaacagcgtcgcaataattatttagtgagagcttttttttttt 898
Db 151 RAKKSAAWTTTWWAARRRRGGGRRRAAAWTTTATWWSGTWTTTWTWTTTKRAAW 92
Qy 899 tataataaaactaaagattttattatttttttttttttttttttttttttttt 940
Db 91 WTTAWKRAAARSMAATTKRAAWKRAWTWTTTTRAWTTWTTGA 50

RESULT 11
CNS0039G/c 1101 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION BACR08K10 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL063921.1 GI:4941778
VERSION 1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 1101)
REFERENCE Direct Submission
AUTHORS Aaron Mammoser in Pieter de Jong's laboratory in the Department of
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's
p1 and Est libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES
    source
        1..1101
            /organism="Drosophila melanogaster"
            /db_xref="taxon:7227"
            /clone_lib="RPCI-98"
            /clone="BACR08K10"
            /note="end : TET3"
BASE COUNT 201 a 64 c 131 g 202 t 503 others
ORIGIN
    Query Match 2.4%; Score 60; DB 13; Length 1101;
    Best Local Similarity 17.3%; Pred. No. 0.03;
    Matches 119; Conservative 269; Mismatches 298; Indels 0; Gaps 0;

Qy 241 gtgaatgtagttgtcagtggaacacattgcttcttcagaaacctccactacagaa 300
Db 1101 GKARRGGDDTWDRTRKDDMDWTWKWTRDADRWRWADGADRWAWDDGAGTWTTATW 1042
Qy 301 aaagctgctacaaacgctatgaatacaataactgtgtttggcagtgctgctcggaag 360
Db 1041 WWWATWTDWTKWKKWWATAAKTDTATWTTWRTAWRADWAGDRGAKRDRDAATDAG 982

```

```

Qy 361 catagtcagggttccttgatcaatgagcagaaaaaacctctatatcttattgagcttcc 420
Db 981 AGRDRGGRRKDKDKDKGDDDDKGGKKKKAACAATAATWTDWDDWDDKWDKWDGAKDRK 922
Qy 421 ttaatgaaccttttagcaaaagctacgaagactagcttaaatggaaaaaacctagct 480
Db 921 ADDDGGAGDKDDGKGKADDDDTGTDKDDDKDKWDDKAGTWDGATWAAATYDWWWW 862
Qy 481 tggttattgaggaaactttgagggaactcagaaaaagctggagactgtgtgccacgta 540
Db 861 GWADADWTTWDAADWADWADWADWADWADWADWADWADWADWADWADWADWADWADW 802
Qy 541 cgttatgagtatgctgaagcctgttcagttccagaaaatagatgtttcaggagattggc 600
Db 801 DRKADRRDAADRRDAAATWTTTTRDTDDMKWKTDTWTRWAADRTWDRDDDDDRD 742
Qy 601 ggtaattattaaagtttggctgcacaaagcaaatgtcttaactatgatcctaaagag 660
Db 741 RAGTAGRWRRRTWKRWRKRRRTWDDADADDTARDRRRRGGDGDADGKGTGKRRRR 682
Qy 661 gctaattgtttacaaatataaaggattttccgcgtctctatatgtatgttggtcacagat 720
Db 681 DRATWDRTDANWADAANWTTTDTDDDKRRRRKRGARRRRRTTARAANWDTWKAWDW 622
Qy 721 tctctatcattcagagcttatgggcttactccaaacagcagcaaacgataaactcggcagt 780
Db 621 AKDWKTRADRWDRWAADTWTDARKARDWAKARAWRARDRRARAAADRRTWTKGTTT 562
Qy 781 gatttactctccgaaagtttgatcaggtatataatttcagcggttttaagtcacaaatttaa 840
Db 561 ATWTTWAARAANWAWAWATTTATWTTTWTWTTTWTWTTTWTWTTTAAWAAWATAT 502
Qy 841 taaaactctttaaagcaggctgcgacttaattatttagtgagagcttttttttttttt 900
Db 501 WAAWTAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 442
Qy 901 taataaaactaaagatttttttttttttttttttttttttttttttttttttttttt 926
Db 441 TTTTWTWTTWAAATATTTTWTWTTTWT 416

RESULT 12
CNS00F0UH/c 996 bp DNA GSS 03-JUN-1999
LOCUS Drosophila melanogaster genome survey sequence TET3 end of BAC:
DEFINITION BACR31021 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL071063.1 GI:4951105
VERSION 1
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
1 (bases 1 to 996)
REFERENCE Direct Submission
AUTHORS Aaron Mammoser in Pieter de Jong's laboratory in the Department of
TITLE Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
JOURNAL NY. The library is named RPCI-98 and was constructed by partial
ECORI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain y2; cn bw sp, the same strain used for the BDGP's

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:22:17 ; Search time 12230.8 Seconds
(without alignments)
878.082 Million cell updates/sec

Title: US-09-391-606-2

Perfect score: 651

Sequence: 1 atgacaaaaaacattatgc.....taggtataatttcagcggttt 651

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb_ba.*
- 2: gb_htg.*
- 3: gb_in.*
- 4: gb_om.*
- 5: gb_ov.*
- 6: gb_pat.*
- 7: gb_ph.*
- 8: gb_pl.*
- 9: gb_pr.*
- 10: gb_ro.*
- 11: gb_sts.*
- 12: gb_sy.*
- 13: gb_un.*
- 14: gb_vl.*
- 15: em_ba.*
- 16: em_fun.*
- 17: em_hum.*
- 18: em_in.*
- 19: em_om.*
- 20: em_or.*
- 21: em_ov.*
- 22: em_pat.*
- 23: em_ph.*
- 24: em_pl.*
- 25: em_ro.*
- 26: em_sts.*
- 27: em_sy.*
- 28: em_un.*
- 29: em_vl.*
- 30: em_htgo_hum.*
- 31: em_htgo_inv.*
- 32: em_htgo_rod.*
- 33: em_htg_hum.*
- 34: em_htg_inv.*
- 35: em_htg_rod.*
- 36: em_htg_other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	647.8	99.5	2238	6	AX045137	Sequence
2	647.8	99.5	10421	1	AE001654	Chlamydia
3	647.8	99.5	12548	1	AE002165	Chlamydia
4	647.8	99.5	300550	1	AP002547	Chlamydia
5	643.8	98.9	2148	1	CHT768DA	Chlamydia
6	240.8	37.0	10954	1	AE001333	Chlamydia
7	237.	36.4	14698	1	AE002357	Chlamydia
8	201.4	30.9	3455	1	CTU50732	Chlamydia
9	46.2	7.1	143364	2	AC015625	Chlamydia
10	44.8	6.9	1141	6	AX083744	Homo sapi
11	41.6	6.4	347550	1	AP001118	Sequence
12	40.4	6.2	174327	9	AC007091	Buchnera
13	39.8	6.1	175384	9	AL451103	Human DNA
14	38.8	6.0	179033	2	AC023295	Homo sapi
15	38.4	5.9	161888	9	HS70501	Homo sapi
16	37.6	5.8	135978	9	CNS01DV8	Human DNA
17	37.6	5.8	151017	9	AL591431	Human chr
18	37.6	5.8	191736	33	AC025411	Human DNA
19	37.6	5.8	193594	2	AL592563	Homo sapi
20	37.4	5.7	41718	3	CELFI6G10	Caenorhab
21	37.4	5.7	62622	2	AC068814	Homo sapi
22	37.4	5.7	144165	9	AC002127	Human BAC
23	37.4	5.7	228999	2	AC083885	Homo sapi
24	37	5.7	159027	2	AC016647	Homo sapi
25	37	5.7	159611	9	AC026475	Homo sapi
26	37	5.7	169195	2	AC090213	Homo sapi
27	37	5.7	185754	2	AP002406	Homo sapi
28	37	5.7	185754	2	AP002406	Homo sapi
29	37	5.7	188072	2	AC009040	Homo sapi
30	36.8	5.7	157000	8	AC007627	Homo sapi
31	36.6	5.6	13933	14	AF013255	Genomic S
32	36.6	5.6	15225	6	AR089138	Human res
33	36.6	5.6	15225	14	AF013254	Sequence
34	36.6	5.6	124321	9	HS417615	Human DNA
35	36.6	5.6	162084	2	AP001093	Homo sapi
36	36.4	5.6	63747	3	AC025727	Caenorhab
37	36.4	5.6	161296	2	AL138963	Homo sapi
38	36.4	5.6	168050	9	AC090526	Homo sapi
39	36.4	5.6	173691	33	AC009823	Homo sapi
40	36.4	5.6	209999	2	AC019128	Homo sapi
41	36.2	5.6	3519	8	ATH297264	Arabidops
42	36.2	5.6	91566	8	F25C20	Arabidops
43	36.2	5.6	110000	2	AC073702_1	Continuation (2 of
44	36.2	5.6	154120	2	AC055114	Homo sapi
45	36.2	5.6	157016	2	AC021866	Homo sapi

ALIGNMENTS

RESULT 1	AX045137	2238 bp	DNA	PAT	24-NOV-2000
LOCUS	Sequence	7 from Patent	WO0066739.		
DEFINITION	AX045137				
ACCESSION	AX045137.1	GI:11343736			
VERSION	AX045137.1				
KEYWORDS	Chlamydia pneumoniae.				
SOURCE	Chlamydia pneumoniae				
ORGANISM	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia				
REFERENCE	1 (bases 1 to 2238)				
AUTHORS	Murkin,A.D., Oomen,R.P., Wang,J. and Dunn,P.				
TITLE	Chlamydia pneumoniae				

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

RESULT	2
AE001654/c	
LOCUS	AE001654 10421 bp DNA
DEFINITION	<i>Chlamydia pneumoniae</i> section 70 of 103 of the complete genome
ACCESSION	AE001654 AE001363
	BCT 01-DEC-2000

```
/codon_start=1
/transl_table=11
/product="ChLPN 76 kDa Homolog_2 (CT623)"
/protein_id="AADI8868.1"
/db_xref="GI:4377034"
/translation="MKQVYQWLASVLLALTISGYAELPSEQKVKSHYTTTLDVEK
DYLKRGFVETRKQDGLVRIAGDVRAWLXYPREDIKNPSDKDKYNPLPNRYRSEFYI
YIDYAEENWLSKNNMTIAGGNTAAGVDINRAFLCYREYKNPEYRTDFPFMVEIGRS
GLGDLFSEQFQSNGLHIYWTRELSKDYQYOVIVHGGPFVVMKTKHVAWVEGI
LNRLPKOFYKCVSVDDNTFVPSSTTEKATNAMKYKVCWQMLKKHSHQVPIWG
QKPLYLYGAPLMPLAKAKTTLNGKENLAWFVGGTGLGLRKAAGDSATVRYEYEA
LSVEIDVSGIGRGNLLKFWFAQAIANIDPEKANGFTNYKGFSAIYMGITDLSFR
AYSGPANDKLGSDFTFRKFDGIISAF"
gene
complement(6516..8159)
/ene="mvln"
/ote="CPn0730"
CDS
complement(6516..8159)
/ene="mvln"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8869.1"
/db_xref="GI:4377035"
/translation="MSRKNEYSLARSFNILLSTGFCRSRTIGFRETAMATYFGADPI
VAATWLGFRVFFLRKLLGLLEQAFIPHEFFURASLDRAAFFRRFRSLIKGSTI
IFTLLIEAVLMVLOYVEEGYEMILLTMILLPCGIFIMMYNVNGLHCENKFFGVG
LAPVYNIWTFYIAARHSDPRERIIGLSVALVIGFFEWLITVPGVWKFLLEAKSP
POEHSVRALLAPLSGLITSIFQLNLLSDICLARYVHEIGPLYLMVSLKIYOLPH
LFGFVFTVLPATSRVOREDHREGKLMKFLVLTLMVMIINTAGLLIALLPVGV
LYEHGLPQSAVAIVRVLYGASIIIPMALAPLVSLVLYAQRQYVPLFGIGITALA
NIVGLTLGRVWLKDVSGISATSIITAWQLYFLWYSSRLPMYSLKLLSIRRSIK
NVTMLACMTLGLNLTQTYVFLNPLTFLAWPLSSITQAIAFLSESCIFLAFI
FGFAKLRLVEDLINLASFEYWRGQRLQRLQHVMDQTON"
8143..8319
/ene="CPn0731"
8143..8319
/ene="CPn0731"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8870.1"
/db_xref="GI:4377036"
/translation="MSFLIKFPFPPPLKQDYLFDFISPTSSSETIGGSYFKLNKASL
QSSRLRSISIIIS"
8654..9535
/ene="nfo"
8654..9535
/ene="CPn0732"
8654..9535
/ene="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AADI8871.1"
/db_xref="GI:4377037"
/translation="MKVLPPSIPLLGAHSTAGLKNALYEGRDIGASTVQIETANO
ROMQRRAKEIVEDFKAALKETDLSYIMSHAGYLINPGADPDVILKSRIGIYQEIIL
CTIGTSFVNFHGAALKSKEDMKNISSFSQSAPLFSDSPPLVLLVETTAGQGT
LIGNSELGYLQNLKNQIPIGVCVDTCHIFAAGYDITSPQWEDVINEFVYGLS
YLAFLHNDNMFPLGANKDRHAPLGEYIGKESPKFLMTDERTKPKYLETPGPGEN
WQKEIGELKFSKNRDS"
complement(9654..10283)
/ene="rs4"
/ote="CPn0733"
complement(9654..10283)
/ene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AADI8872.1"
/db_xref="GI:4377038"
/translation="MARYCGPKNRVARRFGANIFGRSNPLLLKPPHPPGQHGMRKKK
SDYGLQLEOKKLKACYGMIMEKOLVRAKFEVHKGNVQVMAQFLFECEKRLDNVYRM
GFAKTIFAQQQLVAHGHLVNGRRVDRSRFFLRPGMQISLKEKSKRQSVKDALESKD
```

```
BASE COUNT 3278 a 2158 c 1983 g 3002 t
ORIGIN
ESSLPYSVILDKTGFKGELLVSPQDQIEAQLPLPINISVVCEFLSHRT"
Query Match 99.5%; Score 647.8; DB 1; Length 10421;
Best Local Similarity 99.7%; Pred. No. 5.8e-155;
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 1 atgacacacacacattatgcttgggttttagaaggattctcaatcgcttgcctaaacag 60
Db 5798 ATGACAAAAACATTTATGCTTGGGTTGTAGAAAGGATCTCAATCGTTGCCATAACAG 5739
Qy 61 tttttgtgaaatgtagtgtgtgtaactggaacacattcgttccctcagaacccctccat 120
Db 5738 TTTTGTGTAATGTAGTGTGTCGACCTGGACACATTCGTTCCTTCAGAAACCTCCACT 5679
Qy 121 acagaaaagctgtcacaaacgctatgaaatatacaataactgtgttggcagtggtctc 180
Db 5678 ACAGAAAAAGCTGTACAAACGCTATGAATACAAATACTGTGTGTGGCAGTGGCTCGTC 5619
Qy 181 ggaagacatagtcaggttcccttgatcaatgagacagaaagccctctatatctatga 240
Db 5618 GGAAGACATAGTCAGGTTCCTTGGATCAATGGACAGAAAAAGCCCTCTATATCTTTATGGA 5559
Qy 241 gctttcttaataagacccttttagcaaggctcacgaagctacgttaaatggaagaaac 300
Db 5558 GCTTTCTTAATGAACCCCTTAGCAAAAGGCTACGAAAGACTACGTTAAATGGAAACAAAC 5499
Qy 301 ctagtctgtgttattggaggaaactttaggggactcagaaaagctggagactggtctgcc 360
Db 5498 CTAGCTTGGTTTATTGGAGGAACCTTAGGGGACTTCAGAAAAAGCTGGAGACTGGTCTGCC 5439
Qy 361 acagtacgttatgagtatgctgaaccccttgcagcttcagataccagaaatagattttcaggatt 420
Db 5438 ACAGTACGTTATGAGTATGTCGAGCCCTGTGCGGTCCAGAAATAGATGTTTCAGGGATT 5379
Qy 421 ggccgtggtaattattataaagtttgggttcccaagcaaatgtctgctaaactatgacct 480
Db 5378 GGCGCTGCTAATTTATTAAAGTTTGGTTGCGCCCAAGCAATTCGTCTAACTATGATCCT 5319
Qy 481 aaagaggctaataagttttacaaaattataaaggatttccgcctctctatatgtatggc 540
Db 5318 AAAGAGGCTAATGCTGTTTACAAATATATAAGGATTTCCGCTCTATATATGTATGGCATC 5259
Qy 541 acagattctctatcattcagagcttattggcttactccaaacacagcaacgataaactc 600
Db 5258 ACAGATTCTCTATCATTCATTGAGGCTTATGGGCTTACTCAAACACGACAAACGATAAATC 5199
Qy 601 ggccagtgattttacttccgaaagtttgatctagtagtataaatttcagcgttt 651
Db 5198 GGCAGTGATTTTACTTCCGAAAGTTTGTATCTAGGTATATAATTTACAGCGTTT 5148
RESULT 3
AE002165 12548 bp DNA BCT 30-MAY-2000
LOCUS Chlamydia pneumoniae AR39, section 2 of 94 of the complete
DEFINITION genome.
ACCESSION AE002165.1
VERSION AE002165.1
KEYWORDS GI:7188948
SOURCE Chlamydia pneumoniae AR39.
ORGANISM Chlamydia pneumoniae AR39
REFERENCE 1 (bases 1 to 12548)
AUTHORS Read,T.D., Brunham,R.C., Shen,C., Gill,S.R., Heidelberg,J.F.,
White,O., Hickey,E.K., Peterson,J., Umayam,L.A., Utterback,T.,
Berry,K., Bass,S., Linher,K., Weidman,J., Khouri,H., Craven,B.,
Bowman,C., Dodson,R., Gwinn,M., Nelson,W., DeBoy,R., Kolonay,J.,
McClarty,G., Salzberg,S.L., Eisen,J. and Fraser,C.M.
Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
pneumoniae AR39
```

JOURNAL
MEDLINE
PUBMED
REFERENCE
AUTHORS

Nucleic Acids Res. 28 (6), 1397-1406 (2000)

20150255
10684935

2 (bases 1 to 12548)

Read, T.D., Brunham, R.C., Shen, C., Gill, S.R., Heideberg, J.F.,
White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T.,
Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B.,
Bowman, C., Dodson, R., Gwin, M., Nelson, W., DeBoy, R., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.
Direct Submission

Submitted (01-MAR-2000) The Institute for Genomic Research, 9712
Medical Center Dr, Rockville, MD 20850, USA

Location/Qualifiers

1. .12548

/organism="Chlamydia pneumoniae AR39"

/strain="AR39"

/db_xref="taxon:115711"

/note="synonym: Chlamydia pneumoniae AR39"

297. .965

/gene="CP00011"

297. .965

/gene="CP00011"

/note="similar to SP:P31218 GB:X71492 PID:296947 GB:U00096

PID:I736770; identified by sequence similarity; putative"

/codon_start=1

/transl_table=11

/product="uridine kinase"

/protein_id="AAF37907.1"

/db_xref="GI:7188950"

/translation="MLMLMMIIGITGGGAGKTTLTQNIKEIFGEDVSVICODNYK

DRSHYTPERANIWDHDFADNDLISDIKLNNEIVQAPVDFVLGNRSKTEIET

IYPSKVLVEGILVFENQELRDLMDIRIFVTDADERILRMVRDQEGDSVDCIMS

RYLSMVKPMHEKTEPTRKYADIIVHGNVRQNVVNTLSQIKHNLENALESDETYM

VNSK"

complement(1047. .2021)

/gene="CP00012"

complement(1047. .2021)

/gene="CP00012"

/note="conserved hypothetical protein; identified by

Glimmer2; putative"

/codon_start=1

/transl_table=11

/product="conserved hypothetical protein"

/protein_id="AAF37908.1"

/db_xref="GI:7188951"

/translation="MEKKYALAYVITRVNDPHEETALHKKFLEDLDVSCRIVISEQ

GINQFSYSPHAEIYNQWLKERPNFSKIKFKIHKIFPRITVYKRELAALGCE

VDLSQAKHISPOEHEKQENRCLLDVNRNVEKIGFEDNATLPDIQTFREFPEYA

EKLAEQCDPTTVMYCTGIRCELYSPVLEKFEKVYQLDGGVYAYGOQVGTGW

LGLFVFDRLAIPIDESDPDAPVIAECCHCQTPSDAYNCANTDCNALFLCDECIH

QHGCCGECQSQSPRVKFDSSRGKPFRAHLCEISENSASCCLI"

2238. .2867

/gene="CP00013"

2238. .2867

/gene="CP00013"

/note="similar to SP:P81288; identified by sequence

similarity; putative"

/codon_start=1

/transl_table=11

/product="ribosomal protein S4"

/protein_id="AAF37909.1"

/db_xref="GI:7188952"

/translation="MARYCGPKNRRARRFANIGFRSRNPLKKPPHPPGQHGMRKKK

SDYGLQEEKQKACYGMTMEKQVAKFKEVHKQNVACMFLEPFCEKRLDNVYRM

GFATKIFAAQOLVAHGHLVNGREVDRSPFLRPGQISLKEKSKLQSKDALEKSD

ESSLPSVISLDKFGKGLLYSPFQDQIEAQLPLPINISVVCFLSHRT"

complement(2986. .3867)

/gene="CP00014"

complement(2986. .3867)

/gene="CP00014"

/note="similar to SP:P12638 GB:M22591 PID:146954

PID:405898 GB:U00096; identified by sequence similarity;

putative"

/codon_start=1
/transl_table=11
/product="endonuclease IV"
/protein_id="AAF37910.1"
/db_xref="GI:7188953"
/translation="MKVLPSPPLIGAHTSTAGLKNAIYEGRDIGASTVQIIFTAQO
RQWRALKIEVEDKAAKLEDLSYIMSHAGYLINPAGDPDPVILEKSIYQOETL
DCITLIGTSFVPHFPAKALKSKDCMNKIVSSQSAFLDPSPPILVWLETTAGCT
LIGSNFEELGYLVONLKNQIPICGVCDTCHIFAAGYDITSPOGWEDVLNFEDEVGLS
YLAFHINDSMFPLGANKDRHAPLGEYIGKESFKELMDTERTRKIPKYLETTPGGEN
WQKEIGELLKFSKRNDS"
complement(4202. .4378)
/gene="CP00015"
complement(4202. .4378)
/gene="CP00015"
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:7188954"
/translation="MSFLLIKPFSPPLKQDYLFDISPYTSSEITGGSYFKLNKASL
QSTLRLRSISIIIS"
4362. .6005
/gene="CP00016"
4362. .6005
/gene="CP00016"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:7188955"
/translation="MSKQNEVSLARSIFNLSGTCSTRITGIFREJAMATYFGADPI
VAAFGLGRTVFVFLKILGLILEQAFIPHFELRAQSLDRAAFFRRFRLKIGSTI
IFTLLEAVLWLVQVVEGTYDMILLTLLPCGIFLMNVNMGALLHCENKFFGVG
LAPVYNIWIFEVIAARSDPRERIIGLSVALVIGFFFEMLITVPCWKFLEAKSP
POHDSVRALLAPLSGLITLSSIFQLNLLSDICLARYHEIGPLXLYSLKIYOLPH
LFGVFTVLLPAISRCVREDHERGLKMKFVLTLMVSMIITAGLLALLPGLVGV
LIEHGLFPQSAVYAIVRVLRGASIIIPMALAPVLSVLYFAQRYAVPLRIGITGALA
NIVTSLVLRWLVKDSGISYATSIITAWQVLYFYSSKRLPMYKLLMESRRSIK
NMGTLMACMITLGLNLTOTTVIFLNLPLTLPWLPLSSITAOIAIAPLSBSCLFLAPL
FGAKLLRVEDLINLASFVYRGORGLLQHQVMDTON"
6108. .7376
/gene="CP00017"
6108. .7376
/gene="CP00017"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:7188956"
/translation="MKQVQWLASVVVLLATISGYAELPLSEQVKSHYVTLTDEVK
DYLKSGFVETRKODGVLRACVRRARWLYFREDIKNPSDKYNPLPVNRYSEFYL
YIDYRAERNLWSSKNWTAAGGTAAGVDIRNRAFGLGYFKNPTRTDFFMEIGRS
GLGDLFESEVQFSQNFEDGLHIYTRLSKDYPIOVIVHGGFFVYVNMTRKHAYWVEGI
LNRLPKOFVVKQSVDMNTEVPSETTEKATNNAKMYKCYVMQWLVGKHSQYVVEGI
QKPLLYLGAFLMNLAKATKTLNGKENLAWFIGTGLGRKAGDWSATVRYEVEA
LSYPEIDVSGIGRNLKFWFAQAIANYDPKEANGFTNYKGSALYMYGITDLSLSPR
AYCASKPANDKLGSDFTFRKDLGIISAF"
7488. .9443
/gene="CP00018"
7488. .9443
/gene="CP00018"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"


```

ELKDPGPIVIVINSPPGSDVAGFAVWDQIKMLTSPVTVVTVGLAASMGSVLSLCAAP
GRFATPHSNIHOPISGIPITQATDLDIHAREILTKARIIDVYVVEATNQPDII
EKAIDRWMTANEAKDFGLDGLFSFNDL"
complement(1358. .2917)
/gene="glya"
complement(1358. .2917)
/gene="glya"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"
/transl_table="MLKVFKEFKKFAIVEIFTKVAVVSVLLHFKFLENASGKQGSILAS
TAYLAALDHLLNAPPSIGERIIDELKSHRHLKMIASNSVSLSVOLAMNLLTDKYC
EGSPKRFYSCENVDIAEWECVETAKELFAADCACQVPHSGADANLLAYMAILTHKV
QGPVAVKGYKTYNELTEBEYTLIKAMSSCVCLGPSLNSGGHLTHGNVPLNVMKLM
RCFYDNPOTCFDYAEISRLAKYKPKVLIAGTYSRRLNFAVLKQIAEOCGSVL
WDMFAVAGVAGVDEENPIYADIVTTTHTKLRPGGLVLAIRIYESTLNKA
CPLMMGGLPHVIAAKVALKALSDVFKYAHQVNNARRLAERFLSHGLRLTGGT
DNHMMVIDLGSIGSKIAEDILSSVGIAVNRNSLPSDAIGKWDTSIGRIGTALTTL
GMGIDEMEEDVADIIVKVLNRILSCHVEGSSKKNGKELPEAIAQEARDVRVRLLLRP
LYPIDLEALV"
3107. .3778
/gene="CPj0522"
3107. .3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/transl_table="MTLYGLNOKTARKYQAHYLPILTLPYAKSTPONKRALQFLPO
ATHVILSPSTHLFLSRMTLSLTKATLTKTYLICIGESTKERLLSLPGOVKYVVAQO
BIAGIIPQLALPSSARILYPHSSLARPIRELYNRTFFESYPHYTVAPRKLKNI
LSKYKKIIFTSPTVRAFAKIFPRPEKTYWCQGRWTLQEFQFSSQKQVSLLETGK
SRTSP"
3843. .4175
/gene="CPj0523"
3843. .4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/transl_table="MASSATPGDGTAPSLFPPATPRYNEKLFALFYTIALVWIAL
IATIIAGLCHPLCSPIFLTAIPLYFISRYICSHYARNVYIALDVVDPHSHKLDQRS
HSPIFSQR"
4223. .5302
/gene="CPj0524"
4223. .5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/transl_table="MSGSPRTSESSQSVLSVVPDRKETAIPKQFTIAKISTLATIASL
ALGALVAGISITVLGNPFLALITTALESVTVFLYHOMTSKVSSNWQVLEQNEK
PLGKAWQKNDVCYNEMOFYNHLNFKRVAIQTDASQFPQFTFLGLRVEIKNQST
GIIFNPVPTNLNDTATNLTSTLYTLKDKSVWDTCKREGGPAKEDYPPFTEVRV
KLPNEALDQTFNLNLSAEKKSILPTFLGHVCGPKSEELPNQOEYRQALLAYENCL
KAATESHAAIIVALPLFTSVYEVPEEILPKBGTGYWDNQTAFCCKRALLDAIQNTALR
YQSRSLVILDDPNTIESQSRSE"
5642. .6406
/gene="CPj0525"
5642. .6406
/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"

```

```

/transl_table="MHDALLSILAIQELDIKMRMLRVKKEHOKELAKVOSLKSDIRR
KVOEKLEMLNKTQRDGENRIQEISEQINLENQAAVKKMEIDFNAITQEMTTANK
ERRSLQHSLDMLDKOAGGEDILVESLSTENSSQSAVIEKIEFISIKKINEEGKAL
LEORTELKHATNPPELLSIYERLLNNKKDRVVPVPIENRVCSGCHIVLTPQHENVLRKKD
RLIFCHCSRLIYWQESQVNAQENSTAKRRRAAV"
complement(6827. .7816)
/gene="kpsF"
complement(6827. .7816)
/gene="kpsF"
/codon_start=1
/transl_table=11
/product="GutQ/KpsF family sugar-P isomerase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"
/transl_table="MPSPMISTDVCODILKKEAEVDFEFOAFQPKAEQMLAEKILGH
SCHVFSGVGKCVARKVATLOSLSERALFSPVDLLHGLDGLVSPGDIVCLFSKS
GEQELDDVPHLKSRAILVAITSMPYNSLAALSDLVVILPSVAILDPENLPTNST
TCOMIFEDFVHLLHRSVLSSTYKGNHPSQGVQGMKANGKVKDFMPFTEVPFCHLG
DKVSFLEVSFAYCGCVITDPQFLMGIFTDGLRRSLASGYGSEVLSSLSLEKVMTA
NPRCITFEDSDIALQLMRESSPVAVLPLVDNEENRHVTGLLHMHHTLAKAGLL"
complement(7843. .9024)
/gene="sucB_2"
complement(7843. .9024)
/gene="sucB_2"
/codon_start=1
/transl_table=11
/product="dihydrolipoamide succinyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/transl_table="MIFEFPPKIGTSSGGSIVRWLNGDHVANDPELIEVSTKDI
ATLSPKAGRLVFCVNEGEVADGVLGLIELEISEADESTSCPLTSCETKSEA
SSSSSVSPAVLSQNEGIGLDNLQKTAGTGKGRVTRQDLEAYISQOQVSIPE
IFQEVNRIIPMSPLRRAIASLSKSDSDVPHASLVVDVDTDLNMLISGRORFLDTH
GVKLTISFTVQCLAQTLRQFPLNGSLDGTIYMKKSNVGVAVILNKGVVVPIH
NCODRGLTSVAKALADLSSRARLNKLPDSEVQGSVTVTFNFGMIGALIPPIRIPEV
AIIIGIGITQIRVVYVDDDSLAIRKMVYVTLTFDHRVLDGIYSEFITSUKNRLESVTM
G"
complement(9038. .10282)
/gene="glcT"
complement(9038. .10282)
/gene="glcT"
/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/transl_table="MKLWKFIFGLFVGTVLGLVLEDKAIFFKPIGDIFFNLILSMVY
PLVFCSMVIGIASIDMKIGRIGIKSVGLYLGTTALVIGLCFAWIFSPGNCDFPA
QAQSMDSAVTIDSNKTAAYFLSIIAQVFPSPNVRFAEGNIIQIIFIIFLGLIALRL
SGERGRPERFIDGSEIMLRVMYVMSFAPYGVGASMAWISGNHGLVILGKFLII
AYTLACLFLHATLVFGLVRFCKMSFSKFLSSMMDAISCAVSTASSATLPTVMRCVS
KNLGSAEVSQGVFLPLGATVNMNGTAIFOGMAAVFTAAQVNCPLSLSSLLLVVTTAF
SAVGSAGVPGGMITLGSVLASVGLPIQGIATILAGIDRLDVIQVTPMNLGDAVATY
VASGEGELSPYESIKQESVEIT"
complement(10282. .11379)
/gene="ycaH"
complement(10282. .11379)
/gene="ycaH"
/codon_start=1
/transl_table=11
/product="AtPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/transl_table="MKKRFPTFLFYRRVYTIASLEBILGWGLSGLLSKVFAFLVA
CMNFEWSSTPYRARTSVISGNIIVGGAGKPTVIMLAEALRLRGVSCGLVSRGYKSO
SSRQKLTIVVDSKVHVSAYVGDDEPLMAEKLPGSVVWVKDRISARAAREKGIILL
DDGLQYKRLHKHDEIAVVGNDPLGGRAFPGRKLRDFPLRLKTVDAITVNGGKEAG
TVKVRVNAQPIQFVKTIASVWVTHGERIPKRALRELVRGVFCGLGFPQGLNMLRE
EGIHILGKYLPLDHAATTKKELNAYFCQOAMRQOQGLLCTEKDVKSLPRLSGEVSLLP
IAKVEMRLSVNQDDTLLSLNMIQIHNKNGN"

```

Query Match 99.5%; Score 647.8; DB 1; Length 300550;
 Best Local Similarity 99.7%; Pred. No. 6.4e-155;

KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
AUTHORS
TITLE
JOURNAL
FEATURES
Source
gene
CDS
gene
CDS
gene
CDS
gene
CDS
gene
CDS

Chlamydia trachomatis.
Chlamydia trachomatis
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
1 (bases 1 to 10954)
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V. and Davis, R.W.
Genome sequence of an obligate intracellular pathogen of humans:
Chlamydia trachomatis
Science 282 (5389), 754-759 (1998)
99000809
9784136
2 (bases 1 to 10954)
Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hymen, R.W.,
Olinger, L., Grinwood, J., Davis, R.W. and Stephens, R.S.
Comparative genomes of Chlamydia pneumoniae and C. trachomatis
Nat. Genet. 21 (4), 385-389 (1999)
99206606
10192388
3 (bases 1 to 10954)
Stephens, R.S., Kalman, S., Lammel, C.J., Fan, J., Marathe, R.,
Aravind, L., Mitchell, W.P., Olinger, L., Tatusov, R.L., Zhao, Q.,
Koonin, E.V. and Davis, R.W.
Direct Submission
Submitted (20-MAY-1998) Program in Infectious Diseases, University
of California, 235 Warren Hall, Berkeley, CA 94720-7360, USA
Location/Qualifiers
1. .10954
/organism="Chlamydia trachomatis"
/strain="D/UW-3/CX"
/db_xref="taxon:813"
complement(146..2089)
/gene="CT622"
complement(146..2089)
/gene="CT622"
/codon_start=1
/transl_table=11
/product="CHLPN 76kDa Homolog"
/protein_id="AAC68226.1"
/db_xref="GI:3329069"
/translation="MESGPSVSSNQSMNPINQIASNETKESKESPASPSASS
SVSSWFLSSKHALISLRDAIINKNSPTDSLSQLEASTSTVTRVAARDYNEAKS
NFDATAKSLGNATTLAEYETKMDALQMERLAKQAEVTRIKALQKQEVDRK
LNQVLEKQNTKETLTFTTDSADQIPAINSOLEINKNSADQIIKLEGNQISYEAV
LTNAGEVTKASEAGIKLQALQSLVDAGDQSOAQLQAQONNSPDNIAATPKLIDAA
ETKVNELQEHGTLDPLVKKAEQISOAKDQIEIKPSSGSDIPVPGSSAASAGS
AVGALKSSNSGRISLDDVDNEMAAIAMQGFMSIEQFNPNPATAKELQAMEAQL
TAMSDQLGADGELPAETQAIKDALAQKQPSDGLATAMQVAFAAKVGGSAGT
AGTVOMNYKQLYTAFSTSSSSAAALSDGSAYKTLNLSYSESRSQVQSAISQTAN
PALRSVSRSQSGRSADASQRAAETIVRDSOTLGDVYSRLQVLDLSMTIVSNPQ
VNOEIMOKLTASISKAPQFCYPAQVQNSADSLQKFAAQLEFEFVDGERSLAESRENAF
RKQPAFIOQVLNIASLPSGYLS"
complement(2197..3495)
/gene="CT623"
complement(2197..3495)
/gene="CT623"
/codon_start=1
/transl_table=11
/product="CHLPN 76kDa Homolog"
/protein_id="AAC68227.2"
/db_xref="GI:6578109"
/translation="MKFYFYKGFVGGALLAGCSTNLAFQAASSMSDLSVEDLDSYL
SSKGFVETRKGLADGRARWIYAKEDLETQTPAKPMLPTNRYRSFNLYVDY
TAANSWTKNNWVTIAGESSAAGLDINRAFLGVRKYKNPETOAEVFAEIRGLSDG
IFDSVQNSNDGIIHLAARISSEKLPTMIVHGGPFVNNMAKEYAWVEAILNKL
PGNVVYKSVDMNTLTNDPADASAAQPAKPNKYDILVQWLVGKSTAMPFWNG
QTKNLYTGAYLNFLEIPENWQSTPTTKITNGKNHAWFICGLGVRAGDWS
ATVRYEVEALAIPEDVAGIGRQNMKYWFQAQIKQGLDPAKESGNTYKGVSTQYV
MGLTDSVFRAAAYSKPANDNLGSDFTYRKYDLGLISF"
complement(3738..5348)
/gene="mvin"
/note="CT624"

complement(3738..5348)
/gene="mvin"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AAC68228.1"
/db_xref="GI:3329071"
/translation="MRKDDSGSLVRSFLNLLSGTSPSRULTGMLREIVMATYFGADPLV
ASFWLAPRIFFLKLLGSPILGLAFIPHEFLRAQNISRAFFPSFRFFCYSAAIL
FTLIIELGWCSCVSGISFDLTLIIILFSGFLMMYTNSTLLHCKEKFESVGL
APSVNWSIGTVFLARNYDPRNRIFGLAVLIVGFIEMAVTLPGVMKFGOSKEVP
QERDSIRALIALPSLGLSMGIFQLNLLCDMLARYINEVPLYLAYSRIQOLPWHL
FGLGVFTLLPAISRCVQOHOQGYDLIRFSLKLTAVAVVMTMLLILFALPGVRVL
YHGVFPKTAHVAIVEVLGSGSIIPMALAVSALFYARRNYKPYMLVGIIAAVVN
WNLVGLCYCKQVAVLAYATSLVSGOLAMLYCAGKSLPTYKGMWTFKESGKTV
ITTLAAVITIGVNIHTHYVVFIEPLIVPTKPLVSLFDQCGVFAESALFLSVLFG
LAKLKTEDLMNLISFYWKHQHOSILRN"
5521..6387
/gene="nfo"
/note="CT625"
5521..6387
/gene="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AAC68229.1"
/db_xref="GI:3329072"
/translation="MEILPPQEAALLGAHTSAAGLHNLAYEGRDIGATTVOLFTANQ
RQWRRTLTQVMDVDFRIALNETSLYSINSHAGYLNPNPAPNPEILEKTRVCMHQEIA
DCISLGSFVNFHFGAALSRESCLDRTHIASFSQMAPLFENNPPLVLLLETTAGGGS
LIGSFEELAYLQIGKHIPICVCLDTCHIFAAGYDISVAGWESVLYKHFDVIGLS
FLRAIHLSVFLAGKGNKRHAPIGEGICSGDSFCFLMODERTMLPKYLETPGGPDL
WTKIERYLQKVC"
complement(6484..7113)
/gene="rs4"
/note="CT626"
complement(6484..7113)
/gene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AAC68230.1"
/db_xref="GI:3329073"
/translation="MARYCGPKNRIARRFGANIFGRNPLLRKPNPQGHQWRRKK
SDYGLQLEERQKQKACMLLEKQKAYKEVNVKQGNVAQMELEFECRLDNIVYRL
GFAKTIFAQQLVSHGHVLVNGKVRDRSPFVRPMQISLKEKSKRLAIYTESLENKD
OSSLPVLSLDKAAFGELVVAPELDQIASQLPLPVNVSVICFELSHRT"
7497..8480
/gene="ycea"
/note="CT627"
7497..8480
/gene="ycea"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAC68231.1"
/db_xref="GI:3329074"
/translation="MEKNYALAYVYEGVSPNPHREIHALKOLFKTMDVSCRIVISEE
GINGQFSYQOPDAERYNAWLQRPDFAKIFKIHIEENIFPRVTKRYELVALGCS
VDTHKQGHISPEWHEFKLOENRCLVDVRNNYEMWIKGFENAVLPDIETFRFPDYA
DRLAKEHDPKTPVMYCTGIRCELYSALLLEKGFKEYQDGGVLYAGLKMGTGKW
RGKLFVDNRAMPIDEADPNVSPFIARCSLCTDSTIYNCANTCNLFIACCESCIA
THKCCSEECQAPRIAFSAERGNKPFRRKHLCTPTIEQSCCLKEQENQPA"
complement(8552..9427)
/gene="ispa"
/note="CT628"
complement(8552..9427)
/gene="ispa"
/codon_start=1
/transl_table=11
/product="Geranylgeranyl pyrophosphate synthase"
/protein_id="AAC68232.1"
/db_xref="GI:3329075"

Db	2472	GTTTGAAGCTTTACGGATTCCAGAAATTGATGTCGGGTATTGTCGGGAAACCAAAATG	2413
Qy	439	agcttttggcttcgccaagcaaatgtctgctaactatgatcctaaagaggctaaatagtttt	498
Db	2412	AAATATTGGTTTGCTCAAGCTATCAACAAGGATTGGATCCTAAGAATCTTAACGGCGTTT	2353
Qy	499	acaaattataaaggattttccgcctctatatgtatgcatcacagattctctcatcatt	558
Db	2352	ACTAACTATAAAGGAGTTTCCCTATCAGTTGTATTGGCTCTACAGATTTCGTTTCCTTC	2293
Qy	559	agagcttatg99gcttactccaaaacagcaaacgataaactcgccagtgatttcacttc	618
Db	2292	CGAGCTTATGCTGCTTTATTTCTAAGCGTCTCAACGATAACCTTGGTAGGACATTCACCTAT	2233
Qy	619	cgaagcttgatctagggtataaattccagcgtt	650
Db	2232	CGTAAGTATGACCTAGGTTTAATTTCTTCATT	2201

AE002357 1498 bp DNA BCT 26-MAY-2000
Chlamydia muridarum, section 84 of 85 of the complete genome.
 AE002357 AE002160
 AE002357.2 GI:8163344

KEYWORDS
SOURCE ORGANISM
Chlamydia muridarum.
Chlamydia muridarum
Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

ORGANISM

Chlamydia muridarum

Bacteria: Chlamydiales; Chlamydiaceae; Chlamydia.

1 (bases 1 to 14598)

Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F., White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McCarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

REFERENCE

AUTHORS

TITLE. Genome sequences of *Chlamydia trachomatis* MoPn and Chlamydia pneumoniae AR39

JOURNAL Nucleic Acids Res. 28 (6), 1397-1406 (2000)

MEDLINE
PUBMED
REFERENCE
AUTHORS

2 (bases 1 to 14698)
Read, T.D., Brunham, R., Shen, C., Gill, S.R., Heidelberg, J.F.,
White, O., Hickey, E.K., Peterson, J.J., Umayam, L.A., Utterback, B.,
Berry, K.K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, T.,
Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R.J., Kolonay, J.,
McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

White, O., Hickey, E.K., Peterson, J., Umayam, L.A., Utterback, T., Berry, K., Bass, S., Linher, K., Weidman, J., Khouri, H., Craven, B., Bowman, C., Dodson, R., Gwinn, M., Nelson, W., DeBoy, R., Kolonay, J., McClarty, G., Salzberg, S.L., Eisen, J. and Fraser, C.M.

TITLE	Direct Submission
JOURNAL	Submitted (01-MAR-2000) The Institute for Genomic Research, 9712 Medical Center Dr, Rockville, MD 20850, USA
COMMENT	On Jun 1, 2000 this sequence version replaced at-7190933

```

on Jan 17, 2006 this sequence version replaced gi:17090553.
FEATURES
            Location/Qualifiers
                1..14698
                    /organism="Chlamydia muridarum"
                    /ab_xref="taxon:92550"

```

```
/ab_xref=taxon:83560"  
/note=synonym: Chlamydia trachomatis MoPn"  
complement(105..1364)  
/gene=TC0906"
```

```

CDS
=
complement(105..1364)
/gene="TC0906"
/note="conserved hypothetical protein; identified by
Glimmer2; putative"

```

```

/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"

```

```

/db_xref="GI:7190934"
/translation="MTWGFLKQGI FRERSRVGKQFYHCVSRYFHYLPVLAILL
SWFPLSEQQWRYSYFLFPVVSGLWFLAIGLREROLRAAGQLLETKIRKLTERD

```

[illegible]

ידיעה: תאריך: 10.10.2019

Quality coverage: 4.7 in Q20 bases; agarose-fp
Quality coverage: 4.8 in Q20 base.

* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

* 38594 38693: gap of 100 bp
 * 38694 60937: contig of 22244 bp in length
 * 60938 61037: gap of 100 bp
 * 61038 93236: contig of 32199 bp in length
 * 93237 93336: gap of 100 bp
 * 93337 143364: contig of 50028 bp in length.

```
feature
7039..15741
/note="assembly_fragment"
/note="assembly_fragment"
15842..38593
/note="assembly_fragment"
clone_end:sp6
vector_side:right
38694..60937
/note="assembly_fragment"
61038..93236
/note="assembly_fragment"
93337..143364
/note="assembly_fragment"
```

```
ch          7.1%; Score 46.2; DB 2; Length 143364;
l Similarity 51.7%; Pred. No. 0.18;
105: Conservative 0; Mismatches 98; Indels 0;
```

GTGTATATAATTTAGACTACTTAGATATAATATATGCAATTTTGGAAAGAACCTCAAAATAC
 aaaaagctgctctacaaacgctatgaatacaataactggtgttggcagtggtcgtcgcg
 AAACCTGATTCAAGAAAAATAGAAATACTGTATAGATTTCTTACAAAGTAAGATATTCT
 tagcatagtcaggtgtcccttggatcaatggacagaaaaagcctctatatctttatggagag
 TTAGTATAATAAAGTACTTTCCACAAGAAACATCAGGCCCATATTTCTTTATTGAGAG
 tcttaatgaaccccttttagcaaa 266
 TTCTATCAAATACTTAAGTAAA 35039

RESULT 10

/codon_start=1
 /transl_table=11
 /product="ATP synthase C chain"
 /protein_id="BAB12731.1"
 /db_xref="GI:10038696"
 /translation="MENLNVDMLYIAVALMVGSLASIGAAIGILGGKFLGGAARQPD
 LVLRLKRTFFVVMGLVDAIPMIAGVGLYMLFALS"
 3497..3982
 /gene="atpC"
 /note="atpC"
 3497..3982
 /gene="atpC"
 /EC_number="3.6.1.34"
 /note="BU004"
 /codon_start=1
 /transl_table=11
 /product="ATP synthase B chain"
 /protein_id="BAB12732.1"
 /db_xref="GI:10038697"
 /translation="MNLNATILGQAIISFVLFVFCMKYIWPPILAIETROKEIKESL
 TNARKADELYLEKKIHONIIDAKOKASINLANSANKQVSIILEDARNQALESKKII
 NTQSEINAIATHARKNLHKHREVDLSISMAEKIIKKNSKDDNQELLDLVTSLSQV
 N"
 3982..4515
 /gene="atpB"
 3982..4515
 /gene="atpB"
 /EC_number="3.6.1.34"
 /note="BU005"
 /codon_start=1
 /transl_table=11
 /product="ATP synthase delta chain"
 /protein_id="BAB12733.1"
 /db_xref="GI:10038698"
 /translation="MSVADTIARYAOAIFETALNTTIEKWKNLIFIKTIASHKKF
 KNFLSGISPKYLSLIFITIGNIIDENAKNLKLLSNQENILNITFEFEVKLEAC
 YKNIIVQLKSAFLSKENLIRKRVLERFLFKTKIYIKVDPNLLNGMIVKVNITP
 DLSAQHLKQLSDSLNF"
 4530..6068
 /gene="atpA"
 4530..6068
 /gene="atpA"
 /EC_number="3.6.1.34"
 /note="BU006"
 /codon_start=1
 /transl_table=11
 /product="ATP synthase alpha chain"
 /protein_id="BAB12734.1"
 /db_xref="GI:10038699"
 /translation="MRLNSTEISKLKIERIAQFEVFNQSYNGSITISVSDGIKINGL
 SNVMIGEMILLPNNEYATNERTVTGAVMGPYIHLISEGAKVCTOKILEVPVGDN
 FGRVYNALGFDIGKTDNDGFFPVPADAPGVIDRKSVNQPIOTGTGKIDSMIPIG
 RQRELIIGDRTGTALADITIIINQKSGIKCIYVAIGQKLSITINVVKLSEENAL
 LNTIIVASASAAQLYLAIPYSGCMAAEFFRNKGEDSLIIVDDLKSHAVAYRQISLL
 LRTPCPEAFPGDIFYLHSLRLERASRYSMYVQKTKNKTGTGTSITAPIIETGS
 GDVSAPFVNVSITDQGLFESNLNFSGIRPAVNPISVSRVSGAAOTTIKKLSGV
 IRTALAFYDIDKIFKEKILYAHNYFDLMEETKGTDFNIVIKKFKILITEFK
 KNQF"
 6101..6973
 /gene="atpG"
 6101..6973
 /gene="atpG"
 /EC_number="3.6.1.34"
 /note="BU007"
 /codon_start=1
 /transl_table=11
 /product="ATP synthase gamma chain"
 /protein_id="BAB12735.1"
 /db_xref="GI:10038700"
 /translation="MTSTKEIKNKIVSTNTKTKITAMEMVAVSKMRKTEERMRSRP
 YSDIIRKVIDHTQGNLEYKHSLEERKNRIGMIILSTDRGLCGGLNINLKPQVLFP
 IQNFRAVNPICDILFLKSLSVFKLCCGNLAKNGLNENKPLEELINSVGLIIEY
 QCRKDKIFATYANKFNKNSQYPTITQLLPFSKKNDDQASNNNDWLYLIPESKILDT

 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 17 contigs. The true order of the pieces
 * is not known and their order in this sequence record is

```

source
1. .179033
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="RP11-17Pl6"
/clone_lib="RPC1-11 Human Male BAC"
1. 1057
/note="assembly_fragment"
1158. 2610
/note="assembly_fragment"
2711. 5658
/note="assembly_fragment"
5759. 8908
/note="assembly_fragment"
9009. 12605
/note="assembly_fragment"
12706. 15848
/note="assembly_fragment"
clone_end:SP6
vector_side:right
15949. 21456
/note="assembly_fragment"
21557. 27901
/note="assembly_fragment"
28002. 37524
/note="assembly_fragment"
37625. 48049
/note="assembly_fragment"
48150. 59544
/note="assembly_fragment"
59645. 72098
/note="assembly_fragment"
72199. 87255
/note="assembly_fragment"
87356. 104306
/note="assembly_fragment"

```


repeat_region 12586..12667
/note="L1MA8 repeat: matches 6194..6290 of consensus"
L1MA8 repeat: matches 6194..6290 of consensus"
13217..13513
/note="AluSg repeat: matches 1..294 of consensus"
14979..15012
/note="match: GSS: Em:AQ424325"
15211..15965
/note="L1MB3 repeat: matches 5447..6182 of consensus"
15995..16440
/note="MLT1H repeat: matches 29..547 of consensus"
16441..16586
/note="match: GSS: Em:AQ072870"
17144..17261
/note="MLT1C repeat: matches 428..534 of consensus"
18488..18530
/note="Charlie6 repeat: matches 1128..1171 of consensus"
Charlie6 repeat: matches 1128..1171 of consensus"
18534..18760
/note="MER58A repeat: matches 9..216 of consensus"
18795..19024
/note="L2 repeat: matches 2338..2581 of consensus"
19440..19473
/note="L1P repeat: matches 503..534 of consensus"
19862..19959
/note="L2 repeat: matches 2581..2702 of consensus"
20808..20855
/note="L2 copies 4 mer acac 89% conserved"
21699..22143
/note="match: STS: Em:G09929"
21766..21916
/note="MERA repeat: matches 9..189 of consensus"
23787..24263
/note="MLT1C repeat: matches 22..466 of consensus"
24290..24562
/note="LTR16C repeat: matches 117..385 of consensus"
24988..26483
/note="L1M1 repeat: matches -1390..105 of consensus"
28153..28259
/note="MIR repeat: matches 36..176 of consensus"
28986..29049
/note="L6 copies 4 mer cact 79% conserved"
30039..30458
/note="L2 repeat: matches 2112..2519 of consensus"
30691..30931
/note="MIR repeat: matches 39..256 of consensus"
30934..31035
/note="L3 copies 3 mer aat 68% conserved"
31036..31210
/note="match: STS: Em:G09813"
31447..31520
/note="L2 repeat: matches 2632..2705 of consensus"
32120..32216
/note="MLT1H repeat: matches 432..536 of consensus"
32346..32575
/note="MER20 repeat: matches 1..218 of consensus"
32620..32727
/note="L27 copies 4 mer cctt 63% conserved"
32729..33075
/note="L2 repeat: matches 2294..2674 of consensus"
complement(33076..33152)
/note="match: GSS: Em:AQ489056"
33532..33819
/note="AluSg repeat: matches 13..300 of consensus"
33820..33911
/note="L2 repeat: matches 2213..2308 of consensus"
34001..34100
/note="MLT1B repeat: matches 290..390 of consensus"
34261..34326
/note="L3 copies 22 mer 86% conserved"
34267..34326
/note="L30 copies 2 mer tg 88% conserved"
34273..34328

repeat_region 34718..34789
/note="L8 copies 4 mer tqtg 91% conserved"
repeat_region 34722..34783
/note="L31 copies 2 mer tg 85% conserved"
misc_feature 36081..36431
/note="match: GSS: Em:AQ038794"
repeat_region 36566..36720
/note="MER91A repeat: matches 1..156 of consensus"
repeat_region 36723..36972
/note="MIR repeat: matches 7..260 of consensus"
repeat_region 37219..37458
/note="LTR16C repeat: matches 143..386 of consensus"
repeat_region 37931..38230
/note="AluSc repeat: matches 1..301 of consensus"
repeat_region 38237..38310
/note="MIR repeat: matches 47..121 of consensus"
repeat_region 39183..39480
/note="AluSx repeat: matches 2..297 of consensus"
repeat_region 40048..40218
/note="MLT1I repeat: matches 245..410 of consensus"
Query Match 5.9%; Score 38.4; DB 9; Length 161888;
Best Local Similarity 48.2%; Pred. No. 18; Indels 0; Gaps 0;
Matches 108; Conservative 0; Mismatches 116;
QY 19 gcttgggttagaaggattcgaatcgtttgctaaacagttttttgtgaaatgtagt 78
Db 94481 GCCTTGCAGGAGAGAGATTCCACCCCTAGTCCCACTGTGCTACTTCTCTCTG 94422
QY 79 gttgtcactggaacacattcgttccttcgaaacacccactacacagaaagctgctaca 138
Db 94421 GGAGCTGTCCCTGACCCAGTGTGAAACACAGACCCCACTCACCACCCGCAACCAAGA 94362
QY 139 aacgtatgaatacaataactgtttgttcagtggtcgtcggaagcatagtcaggtt 198
Db 94361 TCAATTAGAATTCCTAAAGCATGCTTGACTAGAAACATCAGCAAGAAAGTTCCATT 94302
QY 199 ccttggaatcgaagacagaaaaagcctctatatcttattgagc 242
Db 94301 CCCGTATCTCTAAATGAATAAGCCTCTCTGTCAGTTTGAGC 94258

Search completed: February 7, 2002, 19:27:24
Job time: 13158 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:46:11 ; Search time 685.35 Seconds
(without alignments)
814.356 Million cell updates/sec

Title: US-09-391-606-2

Perfect score: 651

Sequence: 1 atgacaaaaaacattatgc.....taggtataatttcagcgcttt 651

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_1101.*

1:	/SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2:	/SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3:	/SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4:	/SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5:	/SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6:	/SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7:	/SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8:	/SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9:	/SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10:	/SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11:	/SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12:	/SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13:	/SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14:	/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15:	/SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16:	/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17:	/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18:	/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19:	/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20:	/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21:	/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22:	/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	647.8	99.5	1550	21 AAA28411	Chlamydia pneumoni
2	647.8	99.5	2238	21 AAD02066	C. pneumoniae 76 k
3	647.8	99.5	1230025	20 AAX91990	Nucleotide sequenc
C 4	87	13.4	936	22 AAF58252	Oligonucleotide D1
C 5	87	13.4	936	22 AAF58254	Oligonucleotide D1
C 6	87	13.4	936	22 AAF58257	Oligonucleotide D1
C 7	87	13.4	936	22 AAF58259	Oligonucleotide D1
C 8	87	13.4	936	22 AAF58262	Oligonucleotide D2
C 9	87	13.4	938	22 AAF58255	Oligonucleotide D1
10	84.6	13.0	936	22 AAF58252	Oligonucleotide D1
11	84.6	13.0	936	22 AAF58254	Oligonucleotide D1

12	84.6	13.0	936	22 AAF58257	Oligonucleotide D1
13	84.6	13.0	936	22 AAF58259	Oligonucleotide D2
14	84.6	13.0	936	22 AAF58262	Oligonucleotide D2
15	84.6	13.0	938	22 AAF58255	Oligonucleotide D1
C 16	40	6.1	244	22 AAF58238	Oligonucleotide D1
C 17	38.2	5.9	15218	19 AAV18275	RSV isolate 2B wil
C 18	38.2	5.9	15218	20 AAZ22909	Nucleotide sequenc
C 19	38.2	5.9	15218	20 AAZ22914	Nucleotide sequenc
C 20	38.2	5.9	15218	20 AAX35267	DNA encoding the L
C 21	38.2	5.9	15219	19 AAV18277	RSV vaccine 2B33F
C 22	38.2	5.9	15219	19 AAV18278	RSV vaccine 2B20L
C 23	38.2	5.9	15219	19 AAV18279	RSV revertant 2B33
C 24	38.2	5.9	15219	19 AAV18280	RSV revertant 2B20
C 25	38.2	5.9	15219	20 AAZ22911	Nucleotide sequenc
C 26	38.2	5.9	15219	20 AAZ22912	Nucleotide sequenc
C 27	38.2	5.9	15219	20 AAZ22913	DNA encoding the L
C 28	38.2	5.9	15219	20 AAX35269	DNA encoding the L
C 29	38.2	5.9	15219	20 AAX35270	DNA encoding the L
C 30	38.2	5.9	15219	20 AAX35271	DNA encoding the L
C 31	38.2	5.9	15219	20 AAX35272	Oligonucleotide D1
C 32	36.8	5.7	244	22 AAF58238	Respiratory syncyt
C 33	36.6	5.6	15225	19 AAV17552	Respiratory syncyt
C 34	36.6	5.6	15225	21 AAA88744	RSV isolate 18537
C 35	36.6	5.6	15229	19 AAV18276	Nucleotide sequenc
C 36	36.6	5.6	15229	20 AAZ22910	DNA encoding the L
C 37	36.6	5.6	15229	20 AAX35268	DNA encoding the L
C 38	35.6	5.5	2334	18 AAT66428	Canarypox virus C5
C 39	35.6	5.5	2949	13 AAQ25764	Canarypox virus C5
40	35.6	5.5	3208	16 AAT04698	PRW764.5 (contg. C
41	35.6	5.5	3208	17 AAT34124	Canarypox virus Pv
42	35.6	5.5	3208	18 AAT69990	Canarypox virus DN
43	35.6	5.5	3208	20 AAX79930	C5 ORF from canary
44	35.6	5.5	3209	13 AAQ22769	Canarypox PvuII fr
45	35.6	5.5	3209	13 AAQ29107	Canarypox PvuII fr

ALIGNMENTS

RESULT 1

AAA28411	
ID AAA28411 standard; DNA; 1550 BP.	
XX	
AC AAA28411;	
XX	
DT 29-AUG-2000 (first entry)	
XX	
DE Chlamydia pneumoniae lorf2 coding sequence.	
XX	
KW Lorf2; vaccine; antibacterial; antigen; ss.	
XX	
OS Chlamydia pneumoniae.	
XX	
FH Key Location/Qualifiers	
FT CDS 101..1369	
FT	/*tag= a
XX	
PN WO200024901-A1.	
XX	
PD 04-MAY-2000.	
XX	
PF 28-OCT-1999; 99WO-GB03565.	
XX	
PR 28-OCT-1998; 98US-0106037.	
PR 20-SEP-1999; 99US-0154658.	
PR 26-OCT-1999; 99US-0427501.	
XX	
PA (CONN-) CONNAUGHT LAB LTD.	
XX	
PI Murdin AD, Oomen RP, Dunn PL;	
XX	
DR WPI; 2000-350742/30.	
DR P-PSDB; AAY92716.	

XX Isolated polynucleotide encoding a Chlamydia polypeptide useful to
PT treat, diagnose and prevent disease caused by Chlamydia infection
XX
XX
PS Claim 1: Fig 1A-C; 88pp; English.
XX
XX This sequence encodes lorf2 protein of a strain of Chlamydia pneumoniae.
CC Comparison of this sequence as to the recently published genome sequence
CC of C. pneumoniae reveals that the sequence actually contains at least two
CC open reading frames, a first one in the 5' portion and a second one in
CC the 3' portion of the sequence. Despite the presence of the stop codon at
CC the end of this sequence, C. pneumoniae does make a 76 kDa product. It
CC appears possible that C. pneumoniae is able to read through this stop
CC codon and produce a full-length product terminated by the stop codon at
CC the end of the second open reading frame. There is at least one in-frame
CC ATG upstream of the start codon. This suggests that the first open
CC reading frame may form part of one or more larger open reading frames.
CC The lorf2 protein or DNA can be used as a vaccine for humans to treat or
CC prevent disease caused by Chlamydia infection. The sequences or an
CC antibody to lorf2 can be used to diagnose a Chlamydial infection.
XX
SQ Sequence 1550 BP; 472 A; 268 C; 317 G; 493 T; 0 other;

Query Match 99.5%; Score 647.8; DB 21; Length 1550;
Best Local Similarity 99.7%; Pred No. 3.2e-183;
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atgacaaaaaacattatctgttgggttgtagaaggattctcaatcgcttgcctaaacag 60
Db 716 atgacaaaaaacattatctgttgggttgtagaaggattctcaatcgcttgcctaaacag 775
QY 61 tttttgtgaatgtagtgttgcactggaacacattcgttcctcagaaacctccact 120
Db 776 tttttgtgaatgtagtgttgcactggaacacattcgttcctcagaaacctccact 835
QY 121 acagaaaagcgtctacaacgctatgaatacaatacgtgttggcagtggtcgtc 180
Db 836 acagaaaagcgtctacaacgctatgaatacaatacgtgttggcagtggtcgtc 895
QY 181 ggaagcattagtcaggttctctggatcaatggacagaaaagcctctatatctttatgga 240
Db 896 ggaagcattagtcaggttctctggatcaatggacagaaaagcctctatatctttatgga 955
QY 241 gctttcttaataagacctttagcaagccttgcagtaacactacgtttaaaggaaaagaaac 300
Db 956 gctttcttaataagacctttagcaagccttgcagtaacactacgtttaaaggaaaagaaac 1015
QY 301 ctactgtgtttattgaggaacttttaggggactcagaaaagctggagactggtcgc 360
Db 1016 ctactgtgtttattgaggaacttttaggggactcagaaaagctggagactggtcgc 1075
QY 361 acagtacgttatgtagtgcgaagccttgcagttccagaaatagatgtttcaggatt 420
Db 1076 acagtacgttatgtagtgcgaagccttgcagttccagaaatagatgtttcaggatt 1135
QY 421 ggcctgtgtaattattataaagtgttgcgcacaaagcaattgctcctaactatgatcct 480
Db 1136 ggcctgtgtaattattataaagtgttgcgcacaaagcaattgctcctaactatgatcct 1195
QY 481 aaagagcctaagattttacaaaattataaaggattttccgctctatatgtatgcatc 540
Db 1196 aaagagcctaagattttacaaaattataaaggattttccgctctatatgtatgcatc 1255
QY 541 acagattctctattcagagcttatgggcttactcacaacacagcaaacgataaactc 600
Db 1256 acagattctctattcagagcttatgggcttactcacaacacagcaaacgataaactc 1315
QY 601 ggcagtgatttttactttccaaaagtgttgatctaggtataatttcagcgttt 651
Db 1316 ggcagtgatttttactttccaaaagtgttgatctaggtataatttcagcgttt 1366

RESULT 2
AAD02066
ID AAD02066 standard; DNA; 2238 BP.
XX
AC AAD02066;
XX
XX 26-MAR-2001 (first entry)
DT
DE C. pneumoniae 76 kDa protein truncation mutant fusion gene.
XX
XX 76 kDa protein; bactericidal; diagnosis; prevention;
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
KW acute respiratory disease; cough; sore throat; hoarseness; fever;
KW vaccine; immunisation; treatment; truncation mutant; fusion gene;
ds.
XX Chlamydia pneumoniae.
OS Synthetic.
XX
XX Key Location/Qualifiers
FT misc_feature 1..665 /tag= a
FT /note= "This part of the sequence is unrelated to
FT C. pneumoniae 76 kDa gene"
FT CDS 766..2238 /tag= b
FT /product= "Truncated Chlamydia pneumoniae
FT 76kDa protein"
FT misc_feature 2122..2238 /tag= c
FT /note= "This part of the sequence is unrelated to
FT C. pneumoniae 76 kDa gene"
XX
XX WO2000066739-A2.
XX
XX 09-NOV-2000.
PD
XX 03-MAY-2000; 2000WO-CA00511.
XX
XX 03-JUN-1999; 99US-0132270.
PR 30-MAY-1999; 99US-0141276.
XX
XX (AVET) AVENTIS PASTEUR LTD.
XX
XX Murdin AD, Oomen RP, Wang J, Dunn P;
PI
XX WPI: 2000-687542/67.
DR P-PSDB; AAY71957.
XX
XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX useful for vaccinating against Chlamydia infections -
XX
XX Claim 32; Fig 3; 112pp; English.
XX
XX The present sequence is a DNA coding for a fusion protein comprising a
CC truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
CC residues. C. pneumoniae 76 kDa protein is used in the
CC diagnosis, prevention and treatment of C. pneumoniae
CC infections (e.g. pneumonia, upper respiratory
CC tract disease, bronchitis, sinusitis and acute respiratory
CC disease such as cough, sore throat, hoarseness, fever; and
CC abnormal chest sounds on auscultation). C. pneumoniae sequence
CC is also used as vaccines for immunising humans against diseases
CC caused by C. pneumoniae.
XX
XX Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;
SQ

Query Match 99.5%; Score 647.8; DB 21; Length 2238;
Best Local Similarity 99.7%; Pred No. 3.7e-183;
Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
QY 1 atgacaaaaaacattatctgttgggttgtagaaggattctcaatcgcttgcctaaacag 60

Db 1 atgacaaaaaacattatgctgggtgtagaaggattctcaatcgtttgcttaaacag 60
 Qy 61 ttttttgaataatgtagttgtagactggaacacattcgttcttcagaaaccccaact 120
 Db 61 ttttttgaataatgtagttgtagactggaacacattcgttcttcagaaaccccaact 120
 Qy 121 acagaaaaagctgctacaaacgctatgaatacacaataactgtagttgtagactgctgctc 180
 Db 121 acagaaaaagctgctacaaacgctatgaatacacaataactgtagttgtagactgctgctc 180
 Qy 181 ggaagcatagctagggttccttgatcaatggacagaaaaaacgctctatatctttatgga 240
 Db 181 ggaagcatagctagggttccttgatcaatggacagaaaaaacgctctatatctttatgga 240
 Qy 241 gctttcttaataagacccttttagcaaaagctacgaagactaactgtaaaatggaaaaaagaaac 300
 Db 241 gctttcttaataagacccttttagcaaaagctacgaagactaactgtaaaatggaaaaaagaaac 300
 Qy 301 ctactgtggtttattgaggaacttttaggggactcagaaaagctggagactggtctgccc 360
 Db 301 ctactgtggtttattgaggaacttttaggggactcagaaaagctggagactggtctgccc 360
 Qy 361 acagtacgttatgtagtgcgaagccttgtagttccagaaatagatgttttcaggagatt 420
 Db 361 acagtacgttatgtagtgcgaagccttgtagttccagaaatagatgttttcaggagatt 420
 Qy 421 ggcctggttaattattaaagttttagtgcgccaagcaattgctgctaaactatgactcct 480
 Db 421 ggcctggttaattattaaagttttagtgcgccaagcaattgctgctaaactatgactcct 480
 Qy 481 aaagagcctaattgctttacaaataataaaggattttccgctctatatatatgtatggcactc 540
 Db 481 aaagagcctaattgctttacaaataataaaggattttccgctctatatatatgtatggcactc 540
 Qy 541 acagattctctatcattcagagcttatggggttactcccaaacagcaaaacgataaactc 600
 Db 541 acagattctctatcattcagagcttatggggttactcccaaacagcaaaacgataaactc 600
 Qy 601 ggcagtgattttacttcccaaaagttagtctaggtataggtataatttcagcgcttt 651
 Db 601 ggcagtgattttacttcccaaaagttagtctaggtataggtataatttcagcgcttt 651

RESULT 3
 AAX91990/c
 ID AAX91990 standard; DNA: 1230025 BP.
 XX
 AC AAX91990;
 XX
 DT 13-SEP-1999 (first entry)
 XX
 DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
 XX
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope; ss.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO9927105-A2.
 XX
 PD 03-JUN-1999.
 XX
 PF 20-NOV-1998; 98WO-IB01890.
 XX
 PR 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX
 PI Griffiths R;
 XX
 XX

WPI; 1999-357842/30.
 XX
 PT Genome sequence of Chlamydia pneumoniae
 XX
 PS Claim 1; Page 291-611; 1912pp; English.
 XX
 CC The present sequence represents the complete genome of Chlamydia
 CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
 CC respiratory disease such as pneumonia and bronchitis and is thought
 CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
 CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
 CC encoded by the open reading frames of the C. pneumoniae genome (see
 CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
 CC Vectors containing C. pneumoniae nucleotide sequences can also be
 CC used as immunogenic compositions, especially where the vector directs
 CC the expression of a neutralising epitope of C. pneumoniae.
 XX
 SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;

Query Match 99.5%; Score 647.8; DB 20; Length 1230025;
 Best Local Similarity 99.7%; Pred. No. 4.5e-182;
 Matches 649; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 1 atgacaaaaaacattatgctgggtgtagaaggattctcaatcgtttgcttaaacag 60
 Db 829363 ATGACAAAAAACATTATGCTGGTGTGACAGGGATTCTCAATCGTTTCCCTAAACAG 829304
 Qy 61 ttttttgaataatgtagttgtagctggaacacattcgttcttcagaaaccccaact 120
 Db 829303 TTTTGTGTAATGATGTTGTGCTGACTGGAACACATTCTGTTCTCAGAAACCTCCACT 829244
 Qy 121 acagaaaaagctgctacaaacgctatgaatacacaataactgtagttgtagactgctc 180
 Db 829243 ACAGAAAAAGCTGCTACAAACCTATGAATACAAATACTGTGTGTCAGTGGCTCCTC 829184
 Qy 181 ggaagcatagctagggttccttgatcaatggacagaaaaaacgctctatatctttatgga 240
 Db 829183 GGAAGCATAGCTAGGTTCTTGGATCAATGCAGAAAAAGCCTCTATATCTTTATGGA 829124
 Qy 241 gctttcttaataagacccttttagcaaaagctacgaagactaactgtaaaatggaaaaaagaaac 300
 Db 829123 GCTTTCCTTAATGAACCCCTTTTAGCAAGCTACGAAGACTACGTTAAATGGAAGAAAGAAC 829064
 Qy 301 ctactgtggtttattgaggaacttttaggggactcagaaaagctggagactggtctgccc 360
 Db 829063 CTAGCTTGGTTTATTGGAGGAACATTAGGGGGACTCAGAAAAAGCTGGAGACTGCTCTGCC 829004
 Qy 361 acagtacgttatgtagtgcgaagccttgtagttccagaaatagatgttttcaggagatt 420
 Db 829003 ACAGTAGCTTATGATGATGTCGAAAGCCTTGTGCGTTTCCAGAAATAGATGTTTCAGGGATT 828944
 Qy 421 ggcctggttaattattaaagttttagtgcgccaagcaattgctgctaaactatgactcct 480
 Db 828943 GGCCGTGTAAATTATTAAAGTTTGTGTCGCCCAAGCAATGTCTGCTAACTATGATCCT 828884
 Qy 481 aaagagcctaattgctttacaaataataaaggattttccgctctatatatatgtatggcactc 540
 Db 828883 AAAGAGGCTAATGCTTTTACAAATATATAAGGATTTTCCGCTCTATATATGATGTCGATC 828824
 Qy 541 acagattctctatcattcagagcttatggggttactcccaaacagcaaaacgataaactc 600
 Db 828823 ACAGATTCTCTATCATTACAGAGCTTATGGGCTTACTCCAACAGCAAGCAATAAATC 828764
 Qy 601 ggcagtgattttacttcccaaaagttagtctaggtataggtataatttcagcgcttt 651
 Db 828763 GGCAGTGATTTTACTTCCGAAAGTTTGTATAGTATATAATTTCAGCGCTTT 828713

RESULT 4
 AAF58252/c
 ID AAF58252 standard; DNA: 936 BP.
 XX
 XX

AC AAF58252;
 XX 24-APR-2001 (first entry)
 DT
 XX
 DE
 XX
 KW
 XX
 KW
 XX
 OS
 XX
 PN
 XX
 PD
 XX
 PF
 XX
 PR
 XX
 PR
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI
 XX
 DR
 XX
 XX
 PT
 PT
 PT
 PT
 XX
 PS
 XX
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 XX
 SQ

Query Match 13.4%; Score 87; DB 22; Length 936;
 Best Local Similarity 0.9%; Pred. No. 3.9e-16;
 Matches 6; Conservative 386; Mismatches 251; Indels 0; Gaps 0;

QY 1 atgacaaaaaacattatgcttggttagaaggattctcaatcgcttgcctaaacag 60
 DB 771 WWWWWW
 QY 61 tttttgtgaatgtagtgggttagaaggattctcaatcgcttgcctaaacag 120
 DB 711 WWWWWW
 QY 121 acgaaaaagcgtgctacaacgctatgaaatacaatactggttggcagtggtcctc 180
 DB 651 WWWWWW
 QY 181 ggaagcatgtagtgccttgcctggaacagaaagcctctctatcttctgga 240
 DB 591 WWWWWW
 QY 241 gctttcattgaaccttagcaaggctacgaagctacgtctaaatggaagaaac 300
 DB 531 WWWWWW
 QY 301 ctgctgtttatgagaaactttaggggactcagaaagcctggaactggtcgc 360
 DB 471 WWWWWW
 QY 361 acagtcgttatgtagtgcgaagccttgcagttccagaaatagatgtttcaggatt 420
 DB 411 WWWWWW

AC AAF58254/C
 XX ID
 XX AAF58254 standard; DNA; 936 BP.
 AC AAF58254;
 XX
 DT
 XX
 DE
 XX
 XX
 XX
 KW
 XX
 KW
 XX
 OS
 XX
 PN
 XX
 PD
 XX
 PF
 XX
 PR
 XX
 PR
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI
 XX
 DR
 XX
 XX
 PT
 PT
 PT
 PT
 XX
 PS
 XX
 CC
 CC
 CC
 CC
 CC
 CC
 CC
 XX
 SQ

Query Match 13.4%; Score 87; DB 22; Length 936;
 Best Local Similarity 0.9%; Pred. No. 3.9e-16;
 Matches 6; Conservative 386; Mismatches 251; Indels 0; Gaps 0;

QY 1 atgacaaaaaacattatgcttggttagaaggattctcaatcgcttgcctaaacag 60
 DB 771 WWWWWW
 QY 61 tttttgtgaatgtagtgggttagaaggattctcaatcgcttgcctaaacag 120
 DB 711 WWWWWW

QY 421 ggcgtggttaattattaaagtttgggtccccaagcaattgctgctaactatgatcct 480
 DB 351 WWWWWW
 QY 481 aaagaggttaattttttacaaattataaagattttccgctctatatgtatggcaltc 540
 DB 291 WWWWWW
 QY 541 acagattctctatcattcagagcttatggggttactccaacacgacgataaaactc 600
 DB 231 WWWWWW
 QY 601 ggcagtgattttactttccgaaagttagtgcctataattt 643
 DB 171 WWWWWW

RESULT 5
 AAF58254/C
 ID
 XX AAF58254 standard; DNA; 936 BP.
 AC AAF58254;
 XX
 DT
 XX
 DE
 XX
 XX
 XX
 KW
 XX
 KW
 XX
 OS
 XX
 PN
 XX
 PD
 XX
 PF
 XX
 PR
 XX
 PR
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI
 XX
 DR
 XX
 XX
 PT
 PT
 PT
 PT
 XX
 PS
 XX
 CC
 CC
 CC
 CC
 CC
 CC
 XX
 SQ

Query Match 13.4%; Score 87; DB 22; Length 936;
 Best Local Similarity 0.9%; Pred. No. 3.9e-16;
 Matches 6; Conservative 386; Mismatches 251; Indels 0; Gaps 0;

QY 1 atgacaaaaaacattatgcttggttagaaggattctcaatcgcttgcctaaacag 60
 DB 771 WWWWWW
 QY 61 tttttgtgaatgtagtgggttagaaggattctcaatcgcttgcctaaacag 120
 DB 711 WWWWWW

[illegible]

[illegible][illegible]

[illegible]

RESULT 12

AAF58257
ID AAF58257 standard; DNA; 936 BP.

AA AAF58257:

DT 24-APR-2001 (first entry)

AA
DE
Oligonucleotide D1954.

Electron-transfer group; ETM; mismatch; genotyping;
KW
gene expression: ss.
KW

OS Synthetic.

XX
PN WO200107665-A2.

XX
PD 01-FEB-2001.

26-JUL-2000: 2000WO-US20476.

XX 26-JUL-1999: 99US-0145695.

PR 20-JUL-1999; 9903-0143093;
PR 17-MAR-2000; 2000US-0190259;

PA (CLIN-) CLINICAL MICRO SENSORS INC.

XX
PI Umek RM:

AA
DR
WPI; 2001-159728/16.

Nucleic acids containing electron-transfer group, useful as labels in hybridization assays, e.g. for genotyping, allowing repeat analyses on a single surface -

Example 6: Page 127: 159pp: English.

The present invention relates to a composition comprising two nucleic acids each containing an electron-transfer group (ETM) having different redox potentials. The invention is used for electronic detection of nucleic acids, especially of substitutions (mismatches) and single-nucleotide polymorphisms, e.g. for genotyping, monitoring gene expression.

Sequence 936 BP: 5 A: 142 C: 7 G: 6 T: 776 other: XX

Query Match
13.0%; Score 84.6; DB 22; Length 936;

Best Local Similarity 0.9%; Pred. No. 2e-15;
Matches 6; Conservative 384; Mismatches 253; Indels 0; Gaps 0;

[illegible]

RESULT 13

RESULT 13
AAAF58259
ID AAAF58259 standard: DNA: 936 BP.

XX AC AAF58259:

XX
DT 24-APR-2001 (first entry)

XX DE oligonucleotide p2004.

XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression: ss.

XX Synthetic.

XX PN WO200107665-A2

XX
PD 07 - FEB - 2007

XX
26-JUL-2000: 2000WO-IIS20476

XX
26-JUL-1999: 99JJS-0145695

PR 26-JUL-1999; 990US-0145695:
PR 17-MAR-2000; 2000US-0190259:

XX PA (CLIN-) CLINICAL MICRO SENSORS INC.

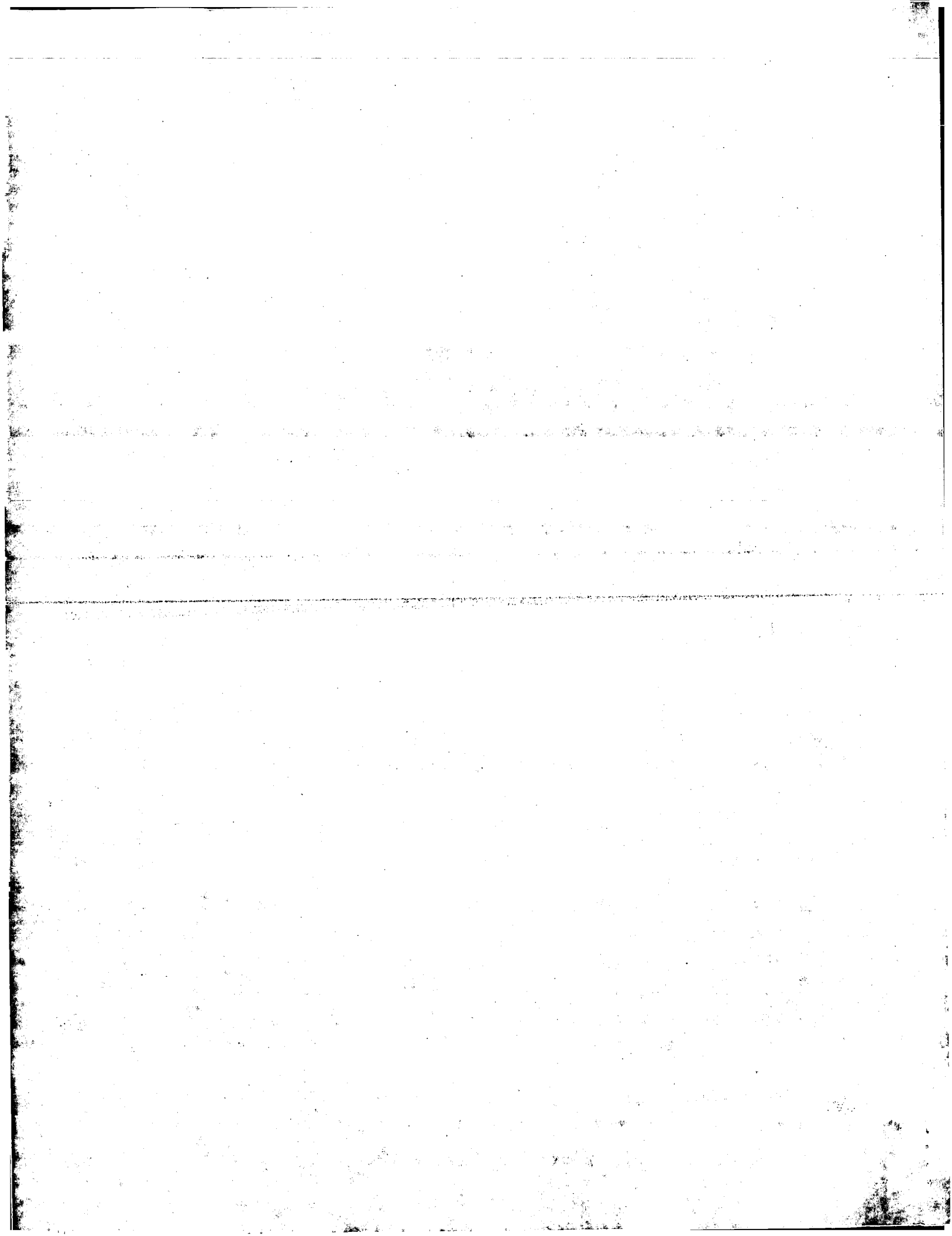
Db 496 555
 QY 421 ggcctgtaatttattaaagtttgggtgcgcacgaattgctgaactatgatcct 480
 Db 556 615
 QY 481 aaagaggctaatagtttacaaattataaaagattttccgctctatatatgatggcgc 540
 Db 616 675
 QY 541 acagattctctatcattcagagcttatgggcttactccaaaccgcaaaacgataaac 600
 Db 676 735
 QY 601 ggcagtgatttactttccgaagtttgatctaggtataattt 643
 Db 736 778

RESULT 15
 AAF58255
 ID AAF58255 standard; DNA; 938 BP.
 AC AAF58255;
 XX 24-APR-2001 (first entry)
 XX Oligonucleotide D1876.
 XX Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 OS Synthetic.
 XX WO200107665-A2.
 XX 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-US20476.
 XX 26-JUL-1999; 99US-0145695.
 XX 17-MAR-2000; 2000US-0190259.
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 XX Umek RM;
 XX WPI; 2001-159728/16.
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX Example 6; Page 127; 159pp; English.
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 13.0%; Score 84.6; DB 22; Length 938;
 Best Local Similarity 0.9%; Pred No. 2e-15;
 Matches 6; Conservative 384; Mismatches 253; Indels 0; Gaps 0;
 QY 1 atgacaaaaaacattatgcttgggtttagaagagattctcaatcgtttgcttaaacag 60
 Db 136 195

Search completed: February 7, 2002, 19:50:20
 Job time: 14534 sec

QY 61 ttttttgtaagttagttgtgcagctggaacacattcgttcttcctcagaacacctccact 120
 Db 196 255
 QY 121 acagaaaaagctgtacaaacgctatgaaatcacaaatagtgtttggcagtggtcgtc 180
 Db 256 315
 QY 181 gaaagcatagtcaggttccttgatcaatggacagaaagcctctctatatcttatgga 240
 Db 316 375
 QY 241 gctttcttaatgaacccottttagcaagctacagagactaagttaattggaagaaac 300
 Db 376 435
 QY 301 ctactgtgtttatttgaggaaactttaggggagctcagaaaagctggagactggtcgc 360
 Db 436 495
 QY 361 acagtacgttatgagtatgtcgaagccttgatccagaaatagatgtttcagggtatt 420
 Db 496 555
 QY 421 ggcctggtaatttattaaagttttggttcgcccaagcaattgctgtaactatgatcct 480
 Db 556 615
 QY 481 aaagaggctaatagtttacaaattataaaagattttccgctctatatatgatggcgc 540
 Db 616 675
 QY 541 acagattctctatcattcagagcttatgggcttactccaaaccgcaaaacgataaac 600
 Db 676 735
 QY 601 ggcagtgatttactttccgaagtttgatctaggtataattt 643
 Db 736 778



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:53:24 ; Search time 243.49 Seconds
(without alignments)
605.516 Million cell updates/sec

Title: US-09-391-606-2

Perfect score: 651

Sequence: 1 atgacaaaaaacattatgc.....taggtataatttcagcgcttt 651

Scoring table: IDENTITY NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 11328999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
c 1	36.6	5.6	15225	2	US-08-892-403A-2
2	35.6	5.5	3208	2	US-08-566-398-51
3	35.6	5.5	3209	1	US-08-105-483-220
4	35.6	5.5	3209	1	US-08-220-151-62
5	35.6	5.5	3209	1	US-08-413-118-62
6	35.6	5.5	3209	1	US-08-224-391-90
7	35.6	5.5	3209	1	US-08-484-304-90
8	35.6	5.5	3209	1	US-08-224-657-39
9	35.6	5.5	3209	1	US-08-709-209-220
10	35.6	5.5	3209	1	US-08-458-101-220
11	35.6	5.5	3209	2	US-08-184-009-68
12	35.6	5.5	3209	2	US-08-486-969-27
13	35.6	5.5	3209	2	US-08-417-210A-66
14	35.6	5.5	3209	2	US-08-458-356-68
15	35.6	5.5	3209	2	US-08-471-025-27
16	35.6	5.5	3209	2	US-08-658-665-27
17	35.6	5.5	3209	3	US-08-473-446-62
18	35.6	5.5	3209	4	US-08-796-101-43
19	35.6	5.5	3209	4	US-08-460-736-68
20	35.6	5.5	3209	4	US-09-085-273-27
21	35.6	5.5	3209	5	PCT-US96-00547-27
22	33.6	5.2	1894	2	US-08-935-450-7
23	32.6	5.0	19011	1	US-08-310-356-36
c 24	32.6	5.0	19557	5	PCT-US92-06300-1
c 25	31.6	4.9	3077	1	US-08-703-809-1
26	31.6	4.9	3077	1	US-08-703-808-1
27	31.6	4.9	3077	2	US-08-914-066-1

28	31.6	4.9	3077	2	US-08-703-807-1	Sequence 1, Appli
29	31.6	4.9	3077	2	US-08-747-108A-1	Sequence 1, Appli
30	31.6	4.9	3077	3	US-09-211-631-1	Sequence 1, Appli
31	31.6	4.9	3077	3	US-09-265-628-1	Sequence 1, Appli
32	31.6	4.9	3077	4	US-09-001-141-1	Sequence 1, Appli
33	31.6	4.9	3077	4	US-09-532-803-1	Sequence 1, Appli
c 34	31	4.8	2433	1	US-08-136-743B-40	Sequence 40, Appli
c 35	31	4.8	2663	1	US-08-136-743B-3	Sequence 3, Appli
c 36	31	4.8	19124	2	US-08-487-826B-13	Sequence 13, Appli
37	30.8	4.7	1749	1	US-08-116-098-1	Sequence 1, Appli
38	30.6	4.7	5319	1	US-08-169-927-1	Sequence 1, Appli
c 39	30.6	4.7	7218	1	US-08-232-463-14	Sequence 14, Appli
40	30.4	4.7	2791	2	US-09-016-000-8	Sequence 8, Appli
41	30.2	4.6	1770	3	US-09-006-636-1	Sequence 1, Appli
42	30.2	4.6	1770	4	US-09-006-632-1	Sequence 1, Appli
43	30.2	4.6	2687	3	US-08-714-918-34	Sequence 34, Appli
44	30.2	4.6	2687	4	US-09-265-315-34	Sequence 34, Appli
45	30.2	4.6	2687	4	US-09-265-315-34	Sequence 34, Appli

ALIGNMENTS

RESULT 1

US-08-892-403A-2/c
Sequence 2, Application US/08892403A
Patent No. 5993824

GENERAL INFORMATION:

APPLICANT: Murphy, Brian R.
APPLICANT: Collins, Peter L.
APPLICANT: Whitehead, Stephen S.
APPLICANT: Bukreyev, Alexander A.
APPLICANT: Juhasz, Katalin
TITLE OF INVENTION: PRODUCTION OF ATTENUATED RESPIRATORY
TITLE OF INVENTION: SYNCTIAL VIRUS VACCINES FROM CLONED NUCLEOTIDE SEQUENCES
NUMBER OF SEQUENCES: 14

CORRESPONDENCE ADDRESS:

ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94111-3834

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/892.403A
FILING DATE: 15-JUL-1997
CLASSIFICATION: 424

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 60/047.634
FILING DATE: 23-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/046.141
FILING DATE: 09-MAY-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/021.773
FILING DATE: 15-JUL-1996

ATTORNEY/AGENT INFORMATION:

NAME: Parmelee, Steven W.
REGISTRATION NUMBER: 31,990
REFERENCE/DOCKET NUMBER: 17634-000510
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-467-9600
TELEFAX: 415-576-0300
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 15225 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-892-403A-2

Query Match 5.6%; Score 36.6; DB 2; Length 15225;
Best Local Similarity 55.0%; Pred. No. 0.13; 59; Indels 0; Gaps 0;
Matches 72; Conservative 0; Mismatches 0;
QY 405 agactgttcaggaggtggtggttaattattataaagtttgggttcgcccgaagcaattgc 464
DB 8397 AGACTGTGAGAGAAATGTTGAGTGCATCAATAAGTTTGTAGGTGCCAATGAATTCA 8338
QY 465 tgcatactatgatctcaagaggctaatagttttacaaaattataaaggattttccgctct 524
DB 8337 TCAAAATATATATTTTAAATAAAGCGTGGTGTATCCAAATGTTATGCTTAAATGTAAT 8278
QY 525 atatatgatg 535
DB 8277 ATGTTTTTAIG 8267

RESULT 2
US-08-566-398-51
Sequence 51, Application US/08566398
Patent No. 5858373
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
TITLE OF INVENTION: RECOMBINANT POXVIRUS - FELINE INFECTIONS
TITLE OF INVENTION: PERITONITIS VIRUS, COMPOSITIONS THEREOF, AND METHODS FOR
TITLE OF INVENTION: MAKING AND USING THEM
NUMBER OF SEQUENCES: 63
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford, P.C.
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/566,398
FILING DATE: 01-DEC-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2880
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 51:
SEQUENCE CHARACTERISTICS:
LENGTH: 3208 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-566-398-51

Query Match 5.5%; Score 35.6; DB 2; Length 3208;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;
QY 348 agactggtctgcacagctacgtatgtatgcgaagccttgcagttccagaataaga 407
DB 2446 AGATAGAGGGCCAAAGTAATACAGGATGTTACGGTCTTCAGCTCTCCATAGAGC 2505

QY 408 tgtttcaggaggtggtggttaattattataaagtttgggttcgcccgaagcaattgctgc 467
DB 2506 TGCTATTTGAGAAACAGG--ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2562
QY 468 taactatgatcctaaagaggttaagttttacaaaattataaaggattttccgctctata 527
DB 2563 AAACCTTTAACTATTGCTAAAGATAATCTTATTAATAAATAATATATCAGCTTTAGTA 2622
QY 528 tatgtatgcacacagattctctatcattcagagcttatgggcttactccaaccagc 587
DB 2623 ATATTAATAATATTAATAAATCTTACTTACTTAATAAATCCAGTGGATATGAACATAATAG 2682
QY 588 aa 589
DB 2683 AA 2684

RESULT 3
US-08-105-483-220
Sequence 220, Application US/08105483
Patent No. 5494807
GENERAL INFORMATION:
APPLICANT: Paoletti, Enzo
TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE
TITLE OF INVENTION: STRAIN
NUMBER OF SEQUENCES: 462
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtis, Morris & Safford
ADDRESSEE: c/o William S. Frommer
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/105,483
FILING DATE: 12-AUG-1993
CLASSIFICATION: 424
PRIOR APPLICATION NUMBER: US 07/847,951
APPLICATION NUMBER: 06-MAR-1992
FILING DATE: 06-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2400
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 220:
SEQUENCE CHARACTERISTICS:
LENGTH: 3209 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-08-105-483-220

Query Match 5.5%; Score 35.6; DB 1; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13; 119; Indels 3; Gaps 1;
Matches 120; Conservative 0; Mismatches 119;
QY 348 agactggtctgcacagctacgtatgtatgcgaagccttgcagttccagaataaga 407
DB 2447 AGATAGAGGGCCAAAGTAATACAGGATGTTTACGGTCTTCAGCTCTCCATAGAGC 2506
QY 408 tgtttcaggaggtggtggttaattattataaagtttgggttcgcccgaagcaattgctgc 467

Db 2507 TCGTATTGGTAGGAACAGG---ATATGATGATAAGCTGTTAATCGCATCATGAGCTGTGATGT 2563
Qy 468 taactatgatcctaaagagcctaataagtttatacaaaattataaagattttccgctctata 527
Db 2564 AACTCTTTAACTATTGCTAAAGATAAATCTATTAAAAAATAATATACGTTTAGTA 2623
Qy 528 tatgtatggcatcacagattctctatctatcattcagagcttatgggcttactccaaccagc 587
Db 2624 ATATTAAAAATATTAATAAAGCTATTACTAATAACTCCAGTGGATGATGAACATAATACG 2683
Qy 588 aa 589
Db 2684 AA 2685

RESULT 4

US-08-220-151-62
; Sequence 62, Application US/08220151
; Patent No. 5529780
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C AND 9D AND USES THEREFOR
; NUMBER OF SEQUENCES: 91
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/220,151
; FILING DATE: 30-MAR-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2540
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066 CURTMS
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3209 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-220-151-62

Query Match 5.5%; Score 35.6; DB 1; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 348 agactggtctgccacagtcacgttatgatgtcgaagccttgcagttccagaaataga 407
Db 2447 AGATAGAGGGGCCAAAGTAATAATACAGGATGTTTACGGTTCTTCAGCTCTCCATAGAGC 2506
Qy 408 tgtttcaggattggccgttgtaatttataaagtttgggttcgcccagaagaattgcgc 467
Db 2507 TCGTATTGGTAGGAACAGG---ATATGATGATAAGCTGTTAATCGATCATGAGCTGTGATGT 2563
Qy 468 taactatgatcctaaagagcctaataagtttatacaaaattataaagattttccgctctata 527

Db 2564 AACTCTTTAACTATTGCTAAAGATAAATCTTTAAAAAATAATATACGTTTAGTA 2623
Qy 528 tatgtatggcatcacagattctctatcattcagagcttatgggcttactccaaccagc 587
Db 2624 ATATTAAAAATATTAATAAAGCTATTACTAATAACTCCAGTGGATGATGAACATAATACG 2683
Qy 588 aa 589
Db 2684 AA 2685

RESULT 5

US-08-413-118-62
; Sequence 62, Application US/08413118
; Patent No. 5688920
; GENERAL INFORMATION:
; APPLICANT: PAOLETTI, ENZO
; TITLE OF INVENTION: NUCLEOTIDE AND AMINO ACID SEQUENCES OF
; TITLE OF INVENTION: CANINE HERPESVIRUS 9B, 9C, AND 9D AND USES THEREFOR
; NUMBER OF SEQUENCES: 128
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CURTIS, MORRIS & SAFFORD, P.C.
; STREET: 530 FIFTH AVENUE, 25TH FLOOR
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 10036
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/413,118
; FILING DATE: 29-MAR-1995
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/220,151
; FILING DATE: 30-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: FROMMER, WILLIAM S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2670
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 62:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3209 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-413-118-62

Query Match 5.5%; Score 35.6; DB 1; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 348 agactggtctgccacagtcacgttatgatgtcgaagccttgcagttccagaaataga 407
Db 2447 AGATAGAGGGGCCAAAGTAATAATACAGGATGTTTACGGTTCTTCAGCTCTCCATAGAGC 2506
Qy 408 tgtttcaggattggccgttgtaatttataaagtttgggttcgcccagaagaattgcgc 467
Db 2507 TCGTATTGGTAGGAACAGG---ATATGATGATAAGCTGTTAATCGATCATGAGCTGTGATGT 2563
Qy 468 taactatgatcctaaagagcctaataagtttatacaaaattataaagattttccgctctata 527
Db 2564 AACTCTTTAACTATTGCTAAAGATAAATCTTTAAAAAATAATATACGTTTAGTA 2623

Db 2624 ATATTAAATATATTAACTCTATTACTAATAACTCCAGTGGATATGAACATAATACG 2683

Db 2624 ATATTAAATATGATTAACTCTATTACTAATAACTCCAGTGGATATGAACATATATACG 2683

RESULT 10

US-08-458-101-220
; Sequence 220, Application US/08458101
; Patent No. 5766599
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Perkus, Marion E.
; APPLICANT: Taylor, Jill
; APPLICANT: Tartaglia, James
; APPLICANT: No. 5766599ton, Elizabeth K.
; APPLICANT: Riviere, Michel
; APPLICANT: de Taisne, Charles
; APPLICANT: Limbach, Keith J.
; APPLICANT: Johnson, Gerard P.
; APPLICANT: Pincus, Steven E.
; APPLICANT: Cox, William I.
; APPLICANT: Audonnet, Jean-Christophe Francis
; APPLICANT: Gettig, Russell Robert
; TITLE OF INVENTION: GENETICALLY ENGINEERED VACCINE
; TITLE OF INVENTION: STRAIN
; NUMBER OF SEQUENCES: 467
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; ADDRESSEE: c/o William S. Frommer
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/458,101
; FILING DATE: 01-JUN-1995
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2740
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; INFORMATION FOR SEQ ID NO: 220:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3209 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-458-101-220

Query Match 5.5%; Score 35.6; DB 1; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 348 agactggtctgccacagtcagtattgtagtgcgaagccttgctcagttccagtaataaga 407
Db 2447 AGATAGAGGGCCAAAGTAATACAGAGGATGTTACGGTTCCTCAGCTCCCATAGAGC 2506
Qy 408 tgtttcaggagattggccgtggtaatttataaagtttggctgcgccaagcaattgctgc 467
Db 2507 TGCTATTGGTAGGAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563
Qy 468 taactatgactcaaacaggctaattttacaaaataataaaggattttccgctctata 527
Db 2564 AAACCTCTTAACTATTCTAAGATATCTTATTAATAAATAATATACGCTTTAGTA 2623
Qy 528 tatgtatggcatcacagattctctatcatcagagcttatgggcttactccaaaccagc 587

Db 2624 ATATTAAATATATTAAATAACTCTATTACTATAACTCCAGTGGATATGAACATAATACG 2683
Qy 588 aa 589
Db 2684 AA 2685

RESULT 11

US-08-184-009-68
; Sequence 68, Application US/08184009
; Patent No. 5833975
; GENERAL INFORMATION:
; APPLICANT: Paoletti, Enzo
; APPLICANT: Tartaglia, James
; APPLICANT: Cox, William I.
; TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
; NUMBER OF SEQUENCES: 217
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Curtis, Morris & Safford
; STREET: 530 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10036

COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/184,009
; FILING DATE: 19-JAN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Frommer, William S.
; REGISTRATION NUMBER: 25,506
; REFERENCE/DOCKET NUMBER: 454310-2530
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 840-3333
; TELEFAX: (212) 840-0712
; TELEX: 425066CURTMS
; INFORMATION FOR SEQ ID NO: 68:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3209 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-08-184-009-68

Query Match 5.5%; Score 35.6; DB 2; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;

Qy 348 agactggtctgccacagtcagtattgtagtgcgaagccttgctcagttccagtaataaga 407
Db 2447 AGATAGAGGGCCAAAGTAATACAGAGGATGTTTACGGTTCCTCAGCTCCCATAGAGC 2506
Qy 408 tgtttcaggagattggccgtggtaatttataaagtttggctgcgccaagcaattgctgc 467
Db 2507 TGCTATTGGTAGGAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563
Qy 468 taactatgactcaaacaggctaattttacaaaataataaaggattttccgctctata 527
Db 2564 AAACCTCTTAACTATTCTAAGATATCTTATTAATAAATAATATACGCTTTAGTA 2623
Qy 528 tatgtatggcatcacagattctctatcatcagagcttatgggcttactccaaaccagc 587
Db 2624 ATATTAAATATATTAAATAACTCTATTACTATAACTCCAGTGGATATGAACATAATACG 2683
Qy 588 aa 589

APPLICANT: Cox, William I.
TITLE OF INVENTION: RECOMBINANT VIRUS IMMUNOTHERAPY
NUMBER OF SEQUENCES: 217
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford
STREET: 530 Fifth Avenue
CITY: New York
STATE: NY
COUNTRY: USA
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/458,356
FILING DATE: 02-JUN-1995
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/184,009
FILING DATE: 19-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2530
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
TELEX: 425066CURTMS
INFORMATION FOR SEQ ID NO: 68:
SEQUENCE CHARACTERISTICS:
LENGTH: 3209 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-458-356-68

Query Match 5.5%; Score 35.6; DB 2; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;
Qy 348 agactggtctccacagtcacgtatgatgtcgaagccttgcagttccagaaataga 407
Db 2447 AGATAGAGGGGCCAAAGTAATACAGGATGTTTACGGTCTTCAGCTCTCCATAGAGC 2506
Qy 408 tgtttcagggttgccggtgtaatttataaagtgttgccgccaagcaattgctgc 467
Db 2507 TGCTATTGTTAGGAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563
Qy 468 taactatgatcctaaaggagcctaagttttacaaattataaaggattttccgctctata 527
Db 2564 AAACCTCTTAACCTATTGCTAAAGATAATCTTATTAAAAAATAATATACGTTTAGTA 2623
Qy 528 tatgtatggcatcacagattctctatcattcagagcttatgggcttactccaaaccagc 587
Db 2624 ATATTAAATATATTAACTACTTATTACTTAATTAACCTCCAGTGGATATGAACATAATACG 2683
Qy 588 aa 589
Db 2684 AA 2685

RESULT 15
US-08-471-025-27
; Sequence 27; Application US/08471025
; Patent No. 5989561
; GENERAL INFORMATION:
; APPLICANT: Paolletti, Enzo
; APPLICANT: Fischer, Laurent
; APPLICANT: Legros, Francois-Xavier

TITLE OF INVENTION: RECOMBINANT POXVIRUS - CALICIVIRUS
TITLE OF INVENTION: [RABBIT HEMORRHAGIC DISEASE VIRUS (RDHV)] COMPOSITIONS AND
TITLE OF INVENTION: USES
NUMBER OF SEQUENCES: 47
CORRESPONDENCE ADDRESS:
ADDRESSEE: Curtiss, Morris & Safford, P.C.
STREET: 530 Fifth Avenue, 25th Floor
CITY: New York
STATE: New York
COUNTRY: United States of America
ZIP: 10036
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/471,025
FILING DATE: 06-JUN-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Frommer, William S.
REGISTRATION NUMBER: 25,506
REFERENCE/DOCKET NUMBER: 454310-2650
TELEPHONE: (212) 840-3333
TELEFAX: (212) 840-0712
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 3209 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
US-08-471-025-27

Query Match 5.5%; Score 35.6; DB 2; Length 3209;
Best Local Similarity 49.6%; Pred. No. 0.13;
Matches 120; Conservative 0; Mismatches 119; Indels 3; Gaps 1;
Qy 348 agactggtctccacagtcacgtatgatgtcgaagccttgcagttccagaaataga 407
Db 2447 AGATAGAGGGGCCAAAGTAATACAGGATGTTTACGGTCTTCAGCTCTCCATAGAGC 2506
Qy 408 tgtttcagggttgccggtgtaatttataaagtgttgccgccaagcaattgctgc 467
Db 2507 TGCTATTGTTAGGAACAGG---ATATGATAAAGCTGTTAATCGATCATGGAGCTGATGT 2563
Qy 468 taactatgatcctaaaggagcctaagttttacaaattataaaggattttccgctctata 527
Db 2564 AAACCTCTTAACCTATTGCTAAAGATAATCTTATTAAAAAATAATATACGTTTAGTA 2623
Qy 528 tatgtatggcatcacagattctctatcattcagagcttatgggcttactccaaaccagc 587
Db 2624 ATATTAAATATATTAACTACTTATTACTTAATAACTCCAGTGGATATGAACATAATACG 2683
Qy 588 aa 589
Db 2684 AA 2685

Search completed: February 7, 2002, 15:53:37
Job time: 331 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:10 ; Search time 7419.31 Seconds
(without alignments)
942.878 Million cell updates/sec

Title: US-09-391-606-2
Perfect score: 651
Sequence: 1 atgacacacacattatgc.....taggtataattcagcgttt 651

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues
Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : EST.*
1: em_estfun:*
2: em_esthum:*
3: em_estin:*
4: em_estom:*
5: em_estpl:*
6: em_estba:*
7: em_estro:*
8: em_estov:*
9: em_htc:*
10: gb_est1:*
11: gb_est2:*
12: gb_htc:*
13: gb_gss:*
14: em_gss_fun:*
15: em_gss_hum:*
16: em_gss_inv:*
17: em_gss_pln:*
18: em_gss_pro:*
19: em_gss_rnd:*
20: em_gss_vrt:*
21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	6.0	474	10	AW569587
2	38.2	5.9	498	13	AQ769224
3	37.8	5.8	1894	12	AK013338
4	37.6	5.8	446	13	AZ933264
5	37.2	5.7	294	10	BB173567
6	37	5.7	303	10	AA184415
7	37	5.7	412	10	AA185888
8	37	5.7	531	13	AZ017765
9	36.8	5.7	780	10	AA563513
10	36.8	5.7	1101	13	CNS017V6
11	36.6	5.6	914	13	CNS0077X
12	36.6	5.6	1101	13	CNS0039G

13	35.4	5.6	1101	13	CNS0106X
14	35.2	5.6	555	13	AZ324640
15	35.2	5.6	846	13	AZ689918
16	36	5.5	694	11	BG440127
17	35.8	5.5	419	10	AU086798
18	35.8	5.5	660	10	AW040929
19	35.8	5.5	871	13	CNS04TMN
20	35.8	5.5	1101	13	CNS00JJR
21	35.6	5.5	375	10	BE185305
22	35.6	5.5	643	13	AZ977355
23	35.4	5.4	484	13	AZ117341
24	35.4	5.4	539	11	BF840086
25	35.2	5.4	333	10	AV206371
26	35.2	5.4	475	13	AQ086655
27	35.2	5.4	539	13	AQ751582
28	35.2	5.4	733	13	AZ945029
29	35	5.4	540	13	AQ514730
30	35	5.4	571	13	AZ066718
31	35	5.4	625	10	AV398733
32	35	5.4	1101	13	CNS00DQU
33	34.8	5.3	272	10	BE080531
34	34.8	5.3	565	13	AZ483477
35	34.8	5.3	713	10	BE661144
36	34.8	5.3	880	13	AZ683973
37	34.8	5.3	899	13	AZ538849
38	34.8	5.3	1101	13	CNS017KX
39	34.6	5.3	501	13	AQ777320
40	34.6	5.3	1086	13	CNS00YXK
41	34.4	5.3	341	10	AI567863
42	34.4	5.3	623	13	AZ872370
43	34.4	5.3	679	13	AZ638237
44	34.4	5.3	772	10	BE314620
45	34.4	5.3	834	13	CNS03EF2

ALIGNMENTS

RESULT	1
AW569587	AW569587
LOCUS	474 bp mRNA
DEFINITION	s188h10.y1 Gm-cl031 Glycine max cDNA clone GENOME SYSTEMS CLONE ID: Gm-cl031-1820 5' similar to SW:TCQP_HUMAN P50990 T-COMPLEX PROTEIN 1, THETA SUBUNIT ;, mRNA sequence.
ACCESSION	AW569587
VERSION	AW569587
KEYWORDS	EST
SOURCE	soybean.
ORGANISM	Glycine max
REFERENCE	Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
AUTHORS	Shoemaker,R., Kelm,P., Vodkin,L., Erpelting,J., Coryell,V., Khanna,A., Bolla,B., Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wyllie,T., Underwood,K., Steptoe,B., Theising,B., Allen,M., Bowers,Y., Person,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston,R. and Willson,R.
TITLE	Public Soybean EST project
JOURNAL	Unpublished (1999)
COMMENT	Contact: Shoemaker R/Public Soybean EST Project Public Soybean EST project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu This clone is available through: Genome Systems, Inc. 4633 World Parkway Circle St. Louis, Missouri 63134 For further information call: (800) 430-0030 or (314) 427-3222 FAX: (888) 919-3324 or (314) 427-3324 or contact: clones@genomesystems.com or

Db	221	T	221	
RESULT	7			
LOCUS	AA185888			
DEFINITION	AA185888 412 bp mRNA EST 17-FEB-1997 m335c07.r1 Soares mouse 3NbMS Mus musculus cDNA clone IMAGE:6230522 5', mRNA sequence.			
ACCESSION	AA185888			
VERSION	AA185888.1	GI:1772287		
KEYWORDS	EST.			
SOURCE	house mouse.			
ORGANISM	Mus musculus			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
AUTHORS	1 (bases 1 to 412) Marra, M., Hillier, L., Allen, M., Bowles, M., Dietrich, N., Dubuque, T., Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilson, R. and Waterston, R.			
TITLE	The WashU-HMI Mouse EST Project			
JOURNAL	Unpublished (1996)			
COMMENT	Contact: Marra M/Mouse EST Project WashU-HMI Mouse EST Project Washington University School 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108			

```

Email: mouseestewartson.wustl.edu
This clone is available royalty-free through LLNL ; contact the
IMAGE Consortium (info@image.llnl.gov) for further information.
MGI:383876

Seq primer: -28M13 rev2 from Amersham
High quality sequence stop: 406.
Location/Qualifiers
1. .412
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:623052"
/clone_lib="Soares mouse 3NbMS"
/sex="male"
/tissue_type="Spleen"
/dev_stage="4 weeks"
/lab_host="DH10B"

/notes="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
was primed with a Not I - oligo(dT) primer [5',
TGTACCAATCTAAGTGGAGCGCGCTCTTTTTTTTTTTTTTTTTTTTTTTTTT
3']; double-stranded cDNA was ligated to Eco RI adaptors
(Pharmacia), digested with Not I and cloned into the Not I
and Eco RI sites of the modified pT73 vector. RNA
provided by Dr. Bertrand Jordan. Library went through
three rounds of normalization, and was constructed by
Bento Soares and M.Ratima Bonaldo."

121 a 65 c 70 g 156 t

BASE COUNT
ORIGIN

```

[illegible]

Db 140 TTTATAAGCATAGCTTGTGGTGAATTTCTTCTTGGTGACAAAGCACACATCAGTACTT 199
QY 370 t 370
Db 200 T 200

RESULT 8
AZ017765 531 bp DNA GSS 25-FEB-2000
LOCUS RPCI-23-265B4.TV RPCI-23 Mus musculus genomic clone RPCI-23-265B4,
DEFINITION DNA sequence.
ACCESSION AZ017765
VERSION AZ017765.1 GI:7093149
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 531)
AUTHORS Zhao,S., Nierman,W., Feldblyum,T., Malek,J., Shatsman,S., Akinret,
B., Levins,M., McGann,S., Tsegaye,G., Geer,K., Krol,M., de Jong,P.
and Fraser,C.M.
TITLE Mouse BAC End Sequences from Library RPCI-23
JOURNAL Unpublished (1999)
COMMENT Other_GSSs: RPCI-23-265B4.TJ
Contact: Shaying Zhao
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: szhao@tigr.org
Clones are derived from the mouse BAC library RPCI-23. For BAC
library availability, please contact Pieter de Jong
(pieter@ejong.med.buffalo.edu). Clones may be purchased from
BACPAC Resources (<http://bacpac.med.buffalo.edu/orderingframe.htm>)
or from Resea ch Genetics (info@resgen.com). BAC end page:
http://www.tigr.org/tdb/bac_ends/mouse/bac_end_intro.html
Seq primer: T7
Plate: 265 row: B column: 4
Class: BAC ends.
Location/Qualifiers
1. 531
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="RPCI-23-265B4"
/clone_lib="RPCI-23"
/sex="Female"
/lab_host="DH10B"
/note="Organ: Kidney/Brain; Vector: pBACE3 6; Site:1;
EcoRI; Site:2: EcoRI; Female C57BL/6J mouse kidney and/or
brain genomic DNA was isolated and partially digested
with a combination of EcoRI and EcoRI Methylase. Size
selected DNA was cloned into the pBACE3.6 vector at the
EcoRI sites. The ligation products were transformed into
DH10B electrocompetent cells (BRL Life Technologies)."
BASE COUNT 171 a 114 c 98 g 146 t 2 others
ORIGIN

Query Match 5.7%; Score 37; DB 13; Length 531;
Best Local Similarity 52.3%; Pred. No. 9.7;
Matches 79; Conservative 0; Mismatches 72; Indels 0; Gaps 0;

QY 26 ttgtagaaggattctcaatcggtttgcctaaacagttttttgtgaaatgtgtgtg 85
Db 36 TTTCAACAGGAACTACCTTCCTTGTGGCTTCTGTACTTGCATATTGTAGCCCTCAC 95
QY 86 actggaacacatcgcttcctcagaaacctccactacagaaaaagctgctacaaacgcta 145

Db 96 ACCAAAGGATTTCTTTTCCAGAGTGTCTCAACAAACACACTGGGTATGATNNTA 155
QY 146 tgaatacaataactgtgttttggcagtgacct 176
Db 156 TACACATAAAAGGTGAAATTTGACAGTGGCT 186

RESULT 9
AA563513 780 bp mRNA EST 18-AUG-1997
LOCUS JE00F007.QF3 PME-SD.96 Schistosoma japonicum cDNA clone SJPF007 5',
DEFINITION mRNA sequence.
ACCESSION AA563513
VERSION AA563513.1 GI:2334978
KEYWORDS EST.
SOURCE Schistosoma japonicum.
ORGANISM Schistosoma japonicum.
REFERENCE Eukaryota; Metazoa; Platyhelminthes; Turbellarian Platyhelminths;
Rhabditophora; Eulcithophora; Revertospermatia; Mediofusata;
Neodermata; Trematoda; Digenea; Strigeidida; Schistosomatoidea;
Schistosomatidae; Schistosoma.
1 (bases 1 to 780)
AUTHORS Fan,J. and Brindley,P.J.
TITLE Expressed sequence tags (ESTs) from a cDNA library of Schistosoma
japonicum (Philippine strain) Egg
JOURNAL Unpublished (1997)
COMMENT Contact: Brindley, P.J.
Molecular Parasitology Unit
Queensland Institute of Medical Research
300 Herston Road, Queensland 4029, Australia
Tel: 61 7 3362 0413
Fax: 61 7 3362 0104
Email: paul@qimr.edu.au
PCR Primers
FORWARD: M13 Forward
BACKWARD: M13 Reverse
Insert length: 1100
Seq primer: T3 Reverse
High quality sequence stop: 780.
Location/Qualifiers
1. 780
/organism="Schistosoma japonicum"
/strain="Philippine strain"
/db_xref="taxon:6182"
/clone="SJPF007"
/clone_lib="PME-SD.96"
/sex="Mixed sex"
/tissue_type="Whole organism"
/dev_stage="Egg"
/lab_host="Rabbit"
/note="Vector: Lambda ZAP-II XR (Stratagene); Site:1:
EcoRI; Site:2: Xho I; Eggs of Schistosoma japonicum
(Philippine strain) were purified from livers of rabbits
about 50 days after infection with cercariae at the
College of Public Health, University of The Philippines,
Manila, The Philippines. Subsequently, at The Queensland
Institute of Medical Research, Brisbane, Australia,
messenger RNA was isolated from lysates of the S.
japonicum eggs by oligo dT chromatography, first strand
cDNA synthesis from the mRNA was primed with an oligo
dT-XhoI primer and synthesized with M-MLV reverse
transcriptase, and second strand synthesis was
accomplished with RNase H and T4 DNA polymerase. The
double stranded cDNA was ligated into the phagemid vector
UNI ZAP XR (Stratagene), and the ligation products
packaged using reagents from Stratagene. The library was
amplified using BB4 strain E. coli bacteria. The library
was constructed by Sharon Day, Rolando Tan, Wilfred Tiu,
and Paul Brindley who consider that it contains sequences
expressed by immature and mature eggs of S. japonicum. The
library is named 'PME-SD.96'." 244 a 128 c 99 g 302 t 7 others
BASE COUNT
ORIGIN

QY	291	aaaagaaaac	ctagctggtt	tattggaagaa	actttagg	gggagctc	gaaagctg	gaga	350
Db	806	AAAAAAKAT	GTCTWTA	AKWATWAT	TTRAGT	TRTTCKY	KGCARAAA	AAAAAAWA	AWR 865
QY	351	ctggtctgc	acagtcgt	tatgat	gtatgc	aagcctt	gtcagtc	cagaaatag	tgt 410
Db	866	ARGGKTK	TMSAAG	AGAGGG	GGGGAT	MAARAC	AGCPRKAA	ATAAGAR	AWAWAAGAAW 925
QY	411	ttcagggat	tcggcgtg	gtaatt	tataa	agttt	gtgttc	gcccaag	caattgctgctaa 470
Db	926	AAAAAAD	DAKBAT	WATAA	AAAAAG	DWATAK	KTWTWTA	CATATRA	VAAATGKGTGTTT 985
QY	471	ctatgatc	ctaaag	gagctaa	gtattt	tacaaa	attataa	aaggattt	tcgcgtctctatat 530

RESULT	12				
CNS0039G/C					
LOCUS					
		CNS0039G	1101 bp	DNA	GSS
					03-JUN-1999

Db	741	RAGTAGRWRWTWKRWRKRRDTRWDADADDDTARDRRRCDDGADAGKGGKTKRKR	682
Qy	487	gctaatagttttacaactataaaggagttttccgcgtctatataatgtatgcacacag	546
Db	681	DRATWTRDAWADAANWTTTDTDDDKRDRRKGARRRRRTTARAADWMTWKAWDW	622
Qy	547	tctctacattcagagcttctgggcttactccaaacagcaaacgataaactcgcgagt	606
Db	621	AKWDKTRADWRDWAADTWDARKADWAKARAWARRDRARAARADRSWTTKGT	562
Qy	607	gattttactttccgaagtttgatctaggataattt	643
Db	561	ATWTTWAARAANWAWAWATTTATWTTTWTWTTT	525

RESULT	13		
CNS0106X			
LOCUS	CNS0106X	1101 bp	DNA
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC		
	BACN03K20 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.		
ACCESSION	AL098595.1	GI:5610206	GSS.
VERSION	AL098595		
KEYWORDS	GSS.		
SOURCE	fruit fly.		
ORGANISM	Plasmid Drosophila melanogaster		
	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;		
	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;		
	Muscomorpha; Ephyridioidea; Drosophilidae; Drosophila.		
REFERENCE	1 (bases 1 to 1101)		
AUTHORS	Genoscope.		
TITLE	Direct Submission		

```

JOURNAL      Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage
              BP 191 91006 EVRY cedex - FRANCE (E-mail : segre@genoscope.cns.fr
COMMENT      - Web : www.genoscope.cns.fr)
              Determination of this BAC-end sequence was carried out as part of a
              collaboration with the European Drosophila Genome Project (EDGP) -
              http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
              library (Dros BAC) was made by Alain Billaud at CEPH (Centre
              d'Etude du Polymorphisme Humain) with funding provided by a MRC
              project grant. The DNA was prepared from embryos by Alain Bucheton
              and Genevieve Payan. It has been constructed in the vector
              pBelobAC11.

FEATURES             Location/Qualifiers
     source            1..1101
                     /organism="Drosophila melanogaster"
                     /plasmid="pBelobAC11"
                     /db_xref="taxon:7227"
                     /clone_lib="DrosBAC"
                     /clone="BACN03K20"
                     /note="end : T7"

BASE COUNT    258 a   107 C   60 g   175 t   501 others

ORIGIN

Query Match          5.6%; Score 36.4; DB 13; Length 1101;
Best Local Similarity 14.5%; Pred. No. 16;
Matches 38; Conservative 125; Mismatches 99; Indels 0; Gaps 0;

Qy 188 atgtagctgttccttgatgcataatggacagaaagaagcctctatatctttatggagtcttct 247
| | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 693 AAATGATCGDKARKAKAKAKATKKARAWDTATWTATAADAAADKGAKDKAKAKAKAD 752

Qy 248 taatgaacccttttagcaaggcttaagaagactacgtttaaatggaaaaaaaacctagctt 307
| | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 753 ARDKRRWDAKDKRKRRAKAKAADAADDAKAADAKADAKADAAADGGRGDGKKRR 812

Qy 308 ggattatggaggaaactttaggggaactcaaaaagactggagactggtctgtccacagtlac 367
| | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 813 AKDRKKKKDKAWGDKKKAKDKAAAKADADAGDAKAKRRRAGDKDKAKADAKAKKA 872

Qy 368 gttatgagtatcgaaagccttgtcagttccagaataatagatgtttcaggagatggcgtg 427
| | | | | : | : | | | | | | | | | | | | | | | | | | | | | | | | | |

```


Db 684 CCATGAATGAACAAAACACCATTTCCATATGGTATTTTCTTAAAAAACCCTCAAGA 743
 Qy 267 ggctcagaagactcacgttaaatggaagaaacacctagcttggtttattggaggaaacttt 326
 Db 744 CACTATTCGAATGCATAACATCAATAGGAATGAAGTAGGTTTATTTAATAAAATTT 803
 Qy 327 agggggactcagaaagctggagac 351
 Db 804 CAGTGCCATATACAAATGGGAGTC 828

Search completed: February 7, 2002, 21:32:15
Job time: 20649 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 7, 2002, 19:27:24 ; Search time 12230.8 Seconds
(without alignments)
1982.766 Million cell updates/sec

Title: US-09-391-606-3

Perfect score: 1470

Sequence: 1 atgttgaatctattgtgc.....accatcatcatcatcat 1470

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1472140 seqs, 8248589755 residues

Total number of hits satisfying chosen parameters: 2944280

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl.*

- 1: gb.ba.*
- 2: gb.htg.*
- 3: gb.in.*
- 4: gb.om.*
- 5: gb.ov.*
- 6: gb.pat.*
- 7: gb.ph.*
- 8: gb.pl.*
- 9: gb.pr.*
- 10: gb.ro.*
- 11: gb.sts.*
- 12: gb.sy.*
- 13: gb.un.*
- 14: gb.vi.*
- 15: em.ba.*
- 16: em.fun.*
- 17: em.hum.*
- 18: em.in.*
- 19: em.om.*
- 20: em.or.*
- 21: em.ov.*
- 22: em.pat.*
- 23: em.ph.*
- 24: em.pl.*
- 25: em.ro.*
- 26: em.sts.*
- 27: em.sy.*
- 28: em.un.*
- 29: em.vi.*
- 30: em.htgo.hum.*
- 31: em.htgo.inv.*
- 32: em.htgo.rod.*
- 33: em.htg.hum.*
- 34: em.htg.inv.*
- 35: em.htg.rod.*
- 36: em.htg.other.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

8

Result No.	Score	Match	Length	DB	ID	Description
1	1460.4	99.3	2238	6	AX045137	AX045137 Sequence
2	1389.4	93.2	2148	1	CHT78KDA	L23921 Chlamydia p
3	1359.4	92.5	2156	6	AX045131	AX045131 Sequence
c 4	1359.4	92.5	10421	1	AE001654	AE001654 Chlamydia
5	1359.4	92.5	12548	1	AE002165	AE002165 Chlamydia
c 6	1359.4	92.5	300550	1	AP002547	AP002547 Chlamydia
7	1347	91.6	1456	6	AX045135	AX045135 Sequence
8	1155.4	78.6	1852	6	AX045133	AX045133 Sequence
9	146.8	10.0	150	6	ARI44067	ARI44067 Sequence
10	65	4.4	816	6	AX105043	AX105043 Sequence
11	65	4.4	1200	6	AX092220	AX092220 Sequence
12	65	4.4	1373	6	AX105044	AX105044 Sequence
13	65	4.4	1620	6	AX105036	AX105036 Sequence
c 14	65	4.4	1620	6	AX105037	AX105037 Sequence
15	64.2	4.4	6849	12	AY029367	AY029367 Synthetic
16	63	4.3	4357	6	AX072889	AX072889 Sequence
17	63	4.3	7108	6	AX001326	AX001326 Sequence
18	63	4.3	7108	6	E36262	E36262 Human semap
19	63	4.3	7475	6	AX001323	AX001323 Sequence
20	63	4.3	7475	6	E36259	E36259 Human semap
21	63	4.3	8192	6	AX001324	AX001324 Sequence
22	63	4.3	8192	6	E36260	E36260 Human semap
c 23	56.6	3.9	10954	1	AE001333	AE001333 Chlamydia
24	56.4	3.8	72	6	AX019231	AX019231 Sequence
c 25	54.4	3.7	7218	6	I66494	I66494 Sequence 14
26	53.6	3.6	1537	6	AX155814	AX155814 Sequence
27	51	3.5	5098	1	MCU51235	U51235 Mycoplasma
28	50.6	3.4	125	6	AR077332	AR077332 Sequence
29	50.6	3.4	2085	6	AR008322	AR008322 Sequence
30	50.6	3.4	2085	6	AR027306	AR027306 Sequence
31	50.6	3.4	2085	6	AR035771	AR035771 Sequence
32	50.6	3.4	2085	6	AR037631	AR037631 Sequence
33	50.6	3.4	2085	6	AR078816	AR078816 Sequence
34	50.6	3.4	2085	6	AR079063	AR079063 Sequence
35	50.6	3.4	2085	6	AR084296	AR084296 Sequence
36	50.6	3.4	2085	6	I16740	I16740 Sequence 1
37	50.6	3.4	2085	6	I70472	I70472 Sequence 1
38	50.6	3.4	2085	6	I92332	I92332 Sequence 1
39	50.6	3.4	2086	1	STRSURPROA	M74122 Streptococc
40	50.4	3.4	217	6	AR083872	AR083872 Sequence
41	50.4	3.4	807	9	HA225093	AJ225093 Homo sapi
42	50.2	3.4	819	9	HA225092	AJ225092 Homo sapi
c 43	50.2	3.4	14698	1	AE002357	AE002357 Chlamydia
44	49.8	3.4	913	6	AR054408	AR054408 Sequence
45	49.8	3.4	913	6	AX080641	AX080641 Sequence

ALIGNMENTS

RESULT 1

AX045137

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

FEATURES

SOURCE

AX045137

Sequence 7 from Patent WO0066739.

AX045137

AX045137.1 GI:11343736

Chlamydia pneumoniae.

Chlamydia pneumoniae

Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.

1 (bases 1 to 2238)

Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.

Chlamydia) antigens and corresponding dna fragments and uses

thereof

Patent: WO 0066739-A 7 09-NOV-2000;

Aventis Pasteur Limited (CA)

Location/Qualifiers

1. 2238

/organism="Chlamydia pneumoniae"

/db_xref="taxon:83558"

CDS	766..2238	/note="unnamed protein product"	
		/codon_start=1	
		/transl_table=1	
		/protein_id="CAC17330.1"	
		/db_xref="GI:11343737"	
		/translation="MVNPGIPGIDETERTPPADLSAQGLEASAAANKSAFAORIAGAE	
		AKPKESKTSVERWSILRSVNALMSLADKLGIASSNSSSTSRSDVSTTATPTP	
		PPPTFDYKTAQATYDTIETSLADIQAALVSLQDAVINIKDTRATDEBETAIAREW	
		ETKNADVKVGAQITELAKYASDNQAILDSGLKTSFDLLQALLOSVAANNKAELL	
		KEMQNDVPVPGKPAIAQSLVDQDTATQIERKDGNARIDAYFAGNAGAVENAKSN	
		NSISNIDSAKQATATQIAEAKKPPDPILOEAEQWVIOAEKDLNKPADGSDV	
		PNPCTTVGSKQOQSSITGSRVSMLLIDENETASILMSGFROWIMHFNTPENDSOAA	
		QOELAAQARAKAGDSDSAAALADAKALEAALGKAGQOQIILNALGQIASAAVYSA	
		GVLPGLVIRARYQAYVEQKLISEDLNSAVDHHHHH"	
BASE COUNT	689 a	459 c	497 g 593 t
ORIGIN			
Query Match	99.3%	Score 1460.4;	DB 6; Length 2238;
Best Local Similarity	99.6%	Pred. No. 0;	
Matches 1464;	Conservative 0;	Mismatches 6;	Indels 0; Gaps 0;
Qy	1	atgggttaatcctattggtccagggtcctatagacgaacacagacacacacccctcccgagat	60
Db	766	ATGGGTTAATCCTATTGTTCCAGGTCTTATAGCGAAACAGAGACGACACCTCCCGCAGAT	825
Qy	61	ctttctgctcaaggattggaagcgatgagcagaataaagatgaggaagctcaagaata	120
Db	826	CTTTCTGCTCAAGGATTGGAGCGAGTCGACGAATAAGATGGCGAAGCTCAAGAATA	885
Qy	121	gcaggtgcggaagctaaagcctaaagaatcgaacccgattctgtagagcgatgagcgc	180
Db	886	GCAGGTGCGGAGCTTAAGCCTTAAGAACTTAAGACCGATTCTGTAGACGGATGGAGCATC	945
Qy	181	ttgcgtttctgcagtgtaatgctctatgagttgagtcgagataagctgggtattgcttctagt	240
Db	946	TTGCGTTCCTGCAAGTAAATGCTCTCATGAGTCTGCGCAGATAAGCTGGGTATTGCTTCTAGT	1005
Qy	241	aaacagctgtcttctactagacagatctcagacgttggaactcaacagcagcagcgcacct	300
Db	1006	AACAGCTGCTCTTCTACTAGCAGATCTGCAGAGTGGGACTCAACAGACAGCCGACCACT	1065
Qy	301	acgcctctcccccacagctctgatattataagactcaagcgcaaacagcttaacgatact	360
Db	1066	ACGCTCTCTCCACCCAGGTTTGATGATTATAGACTCAAGCGCAACAGCTTACGATACT	1125
Qy	361	atcttttaactcaactcactagctagacataaggtgtgttggtgagcctccaggatgct	420
Db	1126	ATCTTTTACCCTCAACTCACTAGCTAGCATACAGGCTGCTTTGGTGAGCCTCCAGGATGCT	1185
Qy	421	gtcactaatataaagatacagcgtctactgatgaggaacgcgaatcgctgcggagtg	480
Db	1186	GTCACTAATATTAAGGATACACCGGCTACTGATGAGGAACCGCAATCGCTCGGAGTGG	1245
Qy	481	gaaactaagaatgccgatgcaattaaagtggcgcgcaaaattacagaattagcgaatat	540
Db	1246	GAAACTAAGAATGCCGATGCAGTTAAAGTTGGCGGCAAAATTACAGAATTACGGAATAT	1305
Qy	541	gtctoggataacaagcagctatttgaactcttaggttaactgaactccttcgacacctta	600
Db	1306	GCTTTCGGATTAAGCAAGCGATTCTTGACTCTTTTAGTTAAACTGACTTCCTTCGACCTCTTA	1365
Qy	601	cgagctgctcttctcaatctagcaacaataacaagcagctgagcttctttaaagag	660
Db	1366	CAGGCTGCTCTTCTCAATCTCTAGCAACAACAAATACAAAGCAGCTGAGCTCTTAAAGAG	1425
Qy	661	atgcaagataaaccagtagtccagggaaacgctgcgaattgtctcaattcttagttgat	720
Db	1426	ATGCAAGATAACCCAGTAGTCCAGGGAACCGCTGCAATTCCTCAATCTTTAGTTGAT	1485
Qy	721	cgacagatgctacagcgacacagatagaaagatggaatgcgattgggatgcata	780

RESULT	2
CHT76KDA	CHT76KDA
LOCUS	2148 bp DNA BCT 17-MAR-1994
DEFINITION	Chlamydia pneumoniae 76 kDa protein gene, complete cds.
ACCESSION	L23921
VERSION	L23921.1 GI:435961
KEYWORDS	76 kDa protein.
SOURCE	Chlamydia pneumoniae DNA.
ORGANISM	Chlamydia pneumoniae
REFERENCE	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
AUTHORS	1. (bases 1 to 2148)
TITLE	Perez-Melgosa, M., Kuo, C.-C. and Campbell, L.
	Isolation and characterization of a gene encoding a Chlamydia
	pneumoniae 76-kilodalton protein containing a species-specific
	epitope. Immun. 62, 880-886 (1994)
JOURNAL	Infect. Immun. 62, 880-886 (1994)
MEDLINE	94156481
FEATURES	Location/Qualifiers
	1..2148
	/organism="Chlamydia pneumoniae"
	/db_xref="taxon:83558"
	1..2148
CDS	

AKPKESKTSVERWSILRSVAVNALMSLADKLGIASSNSSSSRSADVDSTTATPTP
 PPTFDYKTOAGTAYDTIFSTSLADLOAVNINIKDNTDEETAAAEW
 ETKNADVKVCAOITELAKVASDNOAILDLSGLTSPDLQALLOSVANNKAAELL
 KEMQDNVPVCKTPAIAQSLVDOTDATAIOIEKDNARIRDAYFAGQNASCAVENAKSN
 NSISNIDSAKAAATATAKTQAEAKPPDPSPILQEAEPQWVIOEKMILNKPADGSDV
 PNFGTGVGSKQGSSTIGSRVSMLLDDAENETASILMSGFQMIHMFNTPDSQAA
 QOELAAQARAAGADSDSAAALADAKALEAALGKAGQOQGLNALGQTASAVVSA
 GVPPAAASSTIGSSVKOLYKTSKSTGSDYKTOISAGYDAYKSIINDYGRANDATROVI
 NNYSTPALTRSVPARTEARPEKPTQOALARVJSGNSRTLGVDYSOVSALQSVMOIIO
 SNPOANNEERLOKITSAVTQPPQFGPIPVOLSDNSQKFIKLESFAESRSRTAAEIK
 ALSPETNSLFIQOVLNIGSLYSYIQ*

BASE COUNT 672 a 461 c 471 g 552 t
 ORIGIN

Query Match 92.5%; Score 1359.4; DB 6; Length 2156;
 Best Local Similarity 99.5%; Pred. No. 0;
 Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Qy 1 atgggttaacatctattgtccaggctctatagacgaacacagacacacaccccccagat 60
 Db 101 ATGGTTATTCCTATTGTTCCAGGTCCTATAGACGAAACAGACGACACCTCCCGCAGAT 160
 Qy 61 cttctctctcaaggattggagcgagtgagcagcaataaagagtgoggaagctccaagaata 120
 Db 161 CTTTCTGCTCAAGGATTGGAGCGAGTGCAGCAATAAAGAGTGCAGGAAGCTCAAGAATA 220
 Qy 121 gcaggtgcgaagctaaagcctaaagaaatctaaagccgattctgtagacgagtgagcattc 180
 Db 221 GCAGGTGCGGAAGCTAAGCCTAAAGAACTAAGACCGATTCTGTAGCGGATGGAGCATC 280
 Qy 181 ttgcgtctgcagtgaaatgctctcatgagctgtggcagataagctgggtattgctcttagt 240
 Db 281 TTGCGGTTGCAAGTGAATGCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT 340
 Qy 241 aacagctogtcttctactagacatctgcagacgtgagcactgaacacagcagccacact 300
 Db 341 AACAGCTGCTCTTCTACTAGCAGATCTGCAGACGTGGAGTCAACGACACGACCGCAGCT 400
 Qy 301 acgctctctccaccacgctctgattattataagactgaagcgaacagcattacagatact 360
 Db 401 ACGCCCTCCACCCACGCTTGTGATTATTAAGATCAAGCGCAACAGCTTACGATACT 460
 Qy 361 atctttacatcaacatcactagctgacatacaggtgctttgtgagcctccagatgct 420
 Db 461 ATCTTTACCTCAACATCACTACTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCT 520
 Qy 421 gtcactaatataaagatacagcggctactgtagaggaacacgcgaatcgctgcggagtg 480
 Db 521 GTCACATAATAAAGGATACAGCGGCTACTGATGAGGAACCGCAATCGCTCGGAGTGG 580
 Qy 481 gaaactaagaatgccgagtcgaattaaagtggcgcaaatcacagaattagcgaatat 540
 Db 581 GAAACTAAGAATGCCGATGCACTTAAAGTTGGCGGCAATATACAGAAATACGGAATAT 640
 Qy 541 gcttcggaataacaaagcagcttctgactcttaggttaaactgaactctcttcgaccttta 600
 Db 641 GCTTCGGATAACACAGCGATTCTTGACTCTTTAGTAAACTGACTTCCCTTCGACCTCTTA 700
 Qy 601 cagactgctctctccaaatctgtagcaaacataaacaagcagctgagcttctttaaagag 660
 Db 701 CAGGCTGCTCTTCTCCAAATCTAGCAAAATAAACAAGCAGCTGAGCTTCTTAAAGAG 760
 Qy 661 atgcaagataacccagtagtccagggaaacgctgcaattgctcaatctttagttgat 720
 Db 761 ATGCAGATAAACCCAGTAGTCCAGGGAAACGCCCTGCAATTCCTCAATCTTTAGTTGAT 820
 Qy 721 cagacagatgctacagcgacacagatagagaagatgggaatgcgattggggatgcata 780
 Db 821 CAGACAGATGCTACACGACACAGATAGAGAAAGATGGAATGCGATTAGGGATGCAAT 880
 Qy 781 ttgcagggcagacagcctgtagtgctgtagaaaatgctaaaatcttaataacagataagc 840

RESULT 4
 AE001654/c
 LOCUS AE001654 10421 bp DNA BCT 01-DEC-2000
 DEFINITION Chlamydia pneumoniae section 70 of 103 of the complete genome.
 ACCESSION AE001654 AE001363
 VERSION AE001654.1 GI:4377031
 KEYWORDS
 SOURCE Chlamydia pneumoniae CWL029.
 ORGANISM Chlamydia pneumoniae CWL029
 Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
 REFERENCE 1 (bases 1 to 10421)
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hymen, R.W.,
 Olinger, L., Grimwood, J., Davis, R.W. and Stephens, R.S.
 TITLE Comparative genomes of Chlamydia pneumoniae and C. trachomatis
 JOURNAL Nat. Genet. 21 (4), 385-389 (1999)
 MEDLINE 99206606
 PUBMED 10192388
 REFERENCE 2 (bases 1 to 10421)
 AUTHORS Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Olinger, L.,
 Grimwood, J., Davis, R.W. and Stephens, R.S.
 TITLE Direct Submission
 JOURNAL Submitted (01-DEC-1998) Program in Infectious Diseases, University
 of California, 235 Earl Warren Hall, Berkeley, CA 94720, USA
 FEATURES
 source Location/Qualifiers
 1. .10421
 /organism="Chlamydia pneumoniae CWL029"
 /strain="CWL029"
 /db_xref="taxon:115713"
 133..2751
 /gene="CpH0727"

CDS

```

133. .2751
/gene="CpN0727"
/codon_start=1
/transl_table=11
/product="CT619 hypothetical protein"
/protein_id="AADI8866.1"
/db_xref="GI:4377032"
/transl_table="MSTFISQNRLLTIGSGESTRIIKLDHKYSGFDPSPVPAINLEELN
SGIVAAALHLVLDGLAELVASVTEIDGALSTISTVRQLMASYLGTLTLTAEQER
DIVSAALALVLDGLAELVASVTEIDGALSTISTVRQLMASYLGTLTLTAEQER
VFSSTVPSKLEHVKQKAAEIOAKQEEIKAVLEAKGVSTEEAILKEYPDII
AADFKFEIEPLHYRAKVGAPIOEMNENAIQLPPTPAITPDNVNENVMNTLSI
LQADDDAKQAPALGQDEIITLQTLVPLVDKTTFTKAEDLIYTATQPLNTASLK
YLDROAEARKYKTVKQNSIONSETKRVVNNRSMLETOLESMFOAOACNCFVTWIS
QANALNATNKYISAVLTSTSMENYGGLLCLSYMYERLADDEKALFDKSVNEXYLIHI
VGGSWNGHAKMAAQELAEYSILCTAVTSQDQIKAYLQTRGRNEFRATRRFFINIGD
QMYQFANETVPGNLTANGAIQIDPLGFIREAMNVTGVEADIVSNQRLNLEFNTA
ATAHVLQIQIAELQKQADDDPGKASFTEKRFVAANWITSESGLDALISMLNSQ
LPQAEFLKPLIEEFNNLAANLSSLLQITNEFSTTSVYSSLSYLVSQSKQNLFL
AGDYETLLAAAREVYTYRDTARCKQAINLVNGLLOKINSLPGATSAQKQEMNATY
YYQYSLSTNLQTLVLESLLAGLAKWTIQTTSNNKYDKSVPKIESFDDWIPTLALESF
YTGFPNLSATGGLGPLEFTQVQSDQYTSQGTQQLNLQNQMTTIQEWTLVSTSMQ
VLNGILSQLAGAIYSN"
/complement(3078. .5033)
/gene="CpN0728"
/complement(3078. .5033)
/gene="CpN0728"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_1 (CT622)"
/protein_id="AADI8867.1"
/db_xref="GI:4377033"
/transl_table="MVNPIGPPIDETERTPPADLSAQGLEASAANKSAEAOAGTAE
AKRESKTDSEVERSLRSVAVNMLSLADKIGIASNSSSTSRSDVDSPTATAPTP
PPPTFDYKTKQAOYDTIFETSLADIQALVSLQDAPYNIKTATDEBETAIAEW
ETKNADAVKVGQATIELAKTASDNQAILDSQKLTISFDLQALQLQSVANNKAAELL
KEMQNDVPVPGKTPAIQAQSLVDQTDATQIEKQDGNIRDAIFAGQNAAGAVENAKSN
NSIIDSAAKAIATARTQIAEQKFPDPSILQEAQPMYIQAEDKLNKIPADGSDV
PNPQTTVGGSKQSGSISIRVSMLLDAENETASILMSGFRQIMHFNTEPNDSQAA
QOELAAQARAADKDDSAALADAKALFAALGKAGQOQILNALGOLASAVVSA
GVPPAAASSIGSSVKOLYKTSKTSQDYKTOISAGYDAYKSIDNAYGRANDATRDVI
NNVSTPALTRSPARTAEKPEKTDQALARVISGNSRTLDGVYQVSAQOSVQIIO
SNPOANNEERQKLTSAVTPPQGYFYVQLSNDSTQKFIKLESLSFAEGSRTAAEIK
ALSFTNSLFIQQVILNIGSLYSLQ"
/complement(5145. .6413)
/gene="CpN0729"
/complement(5145. .6413)
/gene="CpN0729"
/codon_start=1
/transl_table=11
/product="CHLPN 76 kDa Homolog_2 (CT623)"
/protein_id="AADI8868.1"
/db_xref="GI:4377034"
/transl_table="MKQVQWMLASVVLALTSQYAEPLSEOKVKSHYVTTLDEVK
DYLKRGFEVTRKQDGLVRITAGDVRARWLYFREDIKNPSDKDKNPLPNVNRSEFYL
YIDTRERNLSSKNNMTATAGGNTAAGVDINRAFLGIRYKPNPRTDFEFMEIGRS
GLGDLFESEVQNSFDGLHIYTRFELSXDYPYQVIVHGGPFVNNMTKKHYAWYVEGI
LNRLPKQFVKCSYVDWNTFVPSSTTEKATNAMKYKYCVWQWLVKHSQVWPWING
QKPLLYLGAPLMNPLAKATKTLTLNGENLAWFTGGTLGLLRKAGDWSATVRYEYVEA
LSVPEIDVSGIGRGNLLKFWFAQAIANYDPKEANGFTNYKGFSAALYMYGITDLSFR
AYGAYSXPANDKLGSDFTFRKDLGIISAF"
/complement(6516. .8159)
/gene="mvN"
/notes="CpN0730"
/complement(6516. .8159)
/gene="mvN"
/codon_start=1
/transl_table=11
/product="Integral Membrane Protein"
/protein_id="AADI8869.1"
/db_xref="GI:4377035"
/transl_table="MSRKDNVSLARSIFNLSGTFCSRTIGIPREIATAMATYFADPI
VAAPWLGRTVFFLKLGLGILLEQAFIPHPFEFLRAQSLDRAAFRRFRSLKSGTII

```

gene

```

IETLLIEAVLWVLOVVECTYDMILLTPCGIFELMVMYVNGALLHCNKEFFGVG
LAPVYVNIIEFFVIAARHSDPREIRIGLSVALVIGFFEMLITVPVWKFLLKAKSP
PQEDSVRAALLAPLSGLITSSIFQLNLLSDICLARYVHEIGPLYLMSKLYQULPIH
LFGFVFTVLLPAISRCVREDHERGLMKFVLITMVSIMIITAGLLALLPCVRV
LYEHGLPQSAVAYIVRVLRGYASIIIPMALAPLVLYVFAQROYAVPLFTIGTALA
NIVLSVLGRWLVKDVSGISYATISITAWOLYFLWYSSKRLPMYKSKLLWESIRSIK
VMGTTMLACMTTGLNLTQTTYVIFLNPLTPLAWPLSSITQAQIAFLSESCIFLAPL
FGPAKLLRVDELINLASFEYWRGQGLLQROHVMDQTON"
8143. .8319
/gene="CpN0731"
8143. .8319
/gene="CpN0731"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AADI8870.1"
/db_xref="GI:4377036"
/transl_table="MSFLLIKFPSPPLKQDYLFDISPYTTSSEITIGGSYFKLNKASL
QSSTLRLSRSIIS"
8654. .9535
/gene="nfo"
/notes="CpN0732"
8654. .9535
/gene="nfo"
/codon_start=1
/transl_table=11
/product="Endonuclease IV"
/protein_id="AADI8871.1"
/db_xref="GI:4377037"
/transl_table="MKVLPPPSIPLGAHTSTAGLKNAIYEGRIDGASTVQIIFTANO
QCMQRRALKKEVIEDFAALKETDLSYIMSHAGYLINPGADPDVILKRSRIGIYOEIL
DCTFLGTSFVNFHEDKAAKSKSKMKNKIVSSFSQSLAPDFSPPLVLLRTAGQGT
LIGSNFEELGVLVONLKNQIPGVCVDTCHIFAAGYDITSPQGHEDVLNEDYVGLS
YLRPHLNDSMFPLGANKDRHAPLGEYIGKESFKFLMDTERTKIKYLETGPGPEN
WOKIEGELLKFSKNRDS"
/complement(9654. .10283)
/gene="rs4"
/notes="CpN0733"
/complement(9654. .10283)
/gene="rs4"
/codon_start=1
/transl_table=11
/product="S4 Ribosomal Protein"
/protein_id="AADI8872.1"
/db_xref="GI:4377038"
/transl_table="MARYCGPKNVARREFGANIFGRSRNPLLLKPHPGOHGMORKKK
SDYGLQLEEKQKQKACYGMIMEKQOLVKAKEVIHQGNVAGQFLERECELRDNNVYRM
GPAKTIFAAQOLVAGHILVNGRRVRRRPFPRGMQISLAKESKRLOSQALESKD
ESSLPSYISLDKTFKGELIVSPEDQIEAQLPLPINISVVVCEFLSHRT"

```

CDS

```

BASE COUNT 3278 a 2158 c 1983 g 3002 t
ORIGIN
Query Match 92.5%; Score 1359.4; DB 1; Length 10421;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Oy 1 atgggttaatcctattggtccaggctcctataacagaaacagacacacacctccacagat 60
|||||
Db 5033 ATGTTTAACTCTATTGGTCCAGTCTCTATAGCAGAAACAGACGACACCTCCCCAGAT 4974
|||||
Oy 61 ctttctgctcaaggattggaggcgagtgagcagaaataagagtcggaagctcaagaata 120
|||||
Db 4973 CTTTCTGCTCAAGGATTGGAGGCGAGTGCAGCAATAAGAGTGCAGAGCTCAAGAATA 4914
|||||
Oy 121 ccaggtgcggaagcctaagccctaaagaaacttaagaccgattctgtagagcgatggagcatc 180
|||||
Db 4913 GCAGGTGCGGAAGCTAAGCCCTAAGAAATCTAAGACCGGATCTCTAGAGCGATGGAGCATC 4854
|||||
Oy 181 ttgcgtctgcagtgtaagtctctcatgagctctggcagataagctgggtattgcttctagt 240
|||||
Db 4853 TTGCGTTCTGCAGTCAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT 4794
|||||

```


GINQFSGYEPAELYMQLKERNFNSKIKFIHIRENIPRITVYKRELALGCE
VLSKQKHIPQWHEKLOENRCLILDVRNNYEWKIGHFDNATLPDIOTRERPEYA
EKLAQECDETPPMVYCTGGIRCELSVPLEKEKEVEYOLDGVLAYGOOCTGKW
LGLFVEDRLAIPIDESDPDPAIAECCHQCTPSDAYNCANTDCHALFCCDECIH
QHQCCEGECQSFRVRKFDSSRGNKFRRAHLCEISENSEASCCLI"

gene

CDS

/gene="CP0013"
2238..2867
/note="similar to SP:P81288; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=11
/product="ribosomal protein S4"
/protein_id="AAF37909.1"
/db_xref="GI:7188952"
/translation="MARYCGPKNRVARFRGANIGRSRNPILKKPHPPQGHGMORKKK
SDYGLQLEOKLKACYGMIMEKQLVAFKEVIHKQGNVAQMFLEPFECRLDNMVYRM
CFKATIFAQQQLVAGHILVNGRVRDRRFFLFCMQISLSEKSKRLQSVKDALESKD
ESSLPYSILDTGFKGELLVSPQDQTEAQLPLPINISVVCEFLSHRT"
complement(2986..3867)
/gene="CP0014"
complement(2986..3867)
/note="similar to SP:P12638 GB:M22591 PID:146954
PID:405898 GB:U00096; identified by sequence similarity;
putative"
/codon_start=1
/transl_table=11
/product="endonuclease IV"
/protein_id="AAF37910.1"
/db_xref="GI:7188953"
/translation="MKVLPPPSIPLLGAHTSTAGLKNAIVEGRDIGASTVOIFITANO
ROWORALKREVIDEFAALKETDLSYIMSHAGVILNPGADPDVILEKSRIGIYQEIL
DCITLISFVHFPGCAALKSKEDCMKNKIVSPQSAPLEDSPLVLLLETTAGQGT
LIGNFEELGVLONKQIPIGVCVDTCHIFAAIGDITSPQGWEDVNEFEDEVGLS
LYRAFNDMSFPGLGANKRHAPIGEGYIGKESFKFLMTDERTKRIPKYLETPGPGEN
WQKEIGELLKFSKNRDS"
complement(4202..4378)
/gene="CP0015"
complement(4202..4378)
/note="hypothetical protein; identified by Glimmer2;
putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:7188954"
/translation="MSFLLIKFPSPPLKQDLYDFDISPYTSEITIGGSYFKLNKASL
QSSTLRLSISIIIS"
4362..6005
/gene="CP0016"
4362..6005
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:7188955"
/translation="MSRRDNVLSARSIFNLSGTFCSRITGIFREIATAMATYFGADPI
VAAFWLGRFTVFLKILGLLEQAFIPHFELRAQSLDRAAFRRFRSFLIKGSTI
LFTLLIEAVLVVLYVEEGTYDMILLTMILLPCGIFLMYVNVGALLHCKENKFEVVG
LAPVYVNIHIFVIARHSDPRERIGLSVALVIGFFFEWLIIVPGWKLELAKSP
FQEDSVALLAPLSGLTSSIFQNLNLSIDCLARYVHEIGPIYLMYSLKIYOLPIH
LFEGFVTLVLPALPSRCVQREDHERGLKMFVLTMTSMVMIIMTAGLLALLFEGVRV
LYEHGLFPQSAYVIAVRVIGYGASIIIPMALAPLVSVLIFYAQRQYAVPLFTIGITALA
NIVLSLVGRVAVDVGISYATSIWVOLYELWYSSKRLPMYKSLKLESIRRSIK
VKGTMACMLTGLNLTQTYVIFLNPFLPLANPLSSITQAIAFLSCEIFLAFI
FGPAKLLRVDELINLASFEYWRGQGLLQROHVQMDTQN"
6108..7376

gene

CDS

/gene="CP0015"
4362..6005
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="AAF37911.1"
/db_xref="GI:7188954"
/translation="MSFLLIKFPSPPLKQDLYDFDISPYTSEITIGGSYFKLNKASL
QSSTLRLSISIIIS"
4362..6005
/gene="CP0016"
4362..6005
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37912.1"
/db_xref="GI:7188955"
/translation="MSRRDNVLSARSIFNLSGTFCSRITGIFREIATAMATYFGADPI
VAAFWLGRFTVFLKILGLLEQAFIPHFELRAQSLDRAAFRRFRSFLIKGSTI
LFTLLIEAVLVVLYVEEGTYDMILLTMILLPCGIFLMYVNVGALLHCKENKFEVVG
LAPVYVNIHIFVIARHSDPRERIGLSVALVIGFFFEWLIIVPGWKLELAKSP
FQEDSVALLAPLSGLTSSIFQNLNLSIDCLARYVHEIGPIYLMYSLKIYOLPIH
LFEGFVTLVLPALPSRCVQREDHERGLKMFVLTMTSMVMIIMTAGLLALLFEGVRV
LYEHGLFPQSAYVIAVRVIGYGASIIIPMALAPLVSVLIFYAQRQYAVPLFTIGITALA
NIVLSLVGRVAVDVGISYATSIWVOLYELWYSSKRLPMYKSLKLESIRRSIK
VKGTMACMLTGLNLTQTYVIFLNPFLPLANPLSSITQAIAFLSCEIFLAFI
FGPAKLLRVDELINLASFEYWRGQGLLQROHVQMDTQN"
6108..7376

gene

/gene="CP0017"

CDS

6108..7376

/gene="CP0017"

/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37913.1"
/db_xref="GI:7188956"
/translation="MKQVYQWLASVLLATISGYAELPSEQKVKSHVTITLDEVK
DYLKRGFVETRQKQDVLRIAGDVRARWLYFREDIKNPSDKDKYKPNVNRSEFYV
YIDYABRNMLSKMNTAIAGGNTAAGVYDINRAFLGKRYKPNRTDFEFMIGRS
GLGLFESEVOFQSNDCGLHIYWTRELSKDYPOYVIVHGGPEVVMYTKKHVAWVEGI
LNRLPKQFFVKCVVDWNTFVPSSTTEKATNAMKYKVCWQVLGKHQSQVPIG
OKKPLYLXGAFMLNPLAKATKTTILNGENLAWFICGTLGLGRKAGDSATVRYEVEA
LSVPEIDVSGIGRGNLLKFWFAQIAANYDPEANGFTNYKGSFALYMYGITSLSFR
AYGAYSIPANDKLGSDFTFRKFDLGIISAF"
7488..9443
/gene="CP0018"
7488..9443
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37914.1"
/db_xref="GI:7188957"
/translation="MVNPIGPGPIDETERTPPADLSAQGLEASANKSAEORAGAE
AKPKESKTSVERWSILRSAYNALMSLADKLGIIASSNSSSTSRSDVDTATATPT
PPPTFDYKTOAYDTIFTSTSLADIQAALVSLQDAVTNKDTAADEETAIAAEW
ETKMDNAPVCAQITELAKYASDNOAIIIDLSGKLTSTDLQALLOSVAANNKAELL
KEMQDNPVCKTFAIAGSLVDQDATATQTEKGNALRDAYFAGQASGAVENAKSN
NSTINISAKAIAATATQIAEAKKFPDPSILQEAQWVIAQEDLNKIRAPGSDV
PNPTTGGSKQGGSSISIRVSMLLDDAENETASILMSGFRQIMHMTNPDSQAA
QOELEAQAARAGDDSAALADAQKALEALGKAGQOQGIINALGQIASAAVVA
GVPPAAASSIGSSVKQLYKTSKTSQDYKTOISAGYDAYKSIDNAYGRANDATRDVI
NNVSTPALTRSVPRARTEARGPKTDQALARVISGNSRTLDGTVSOVSALQSVMOIIO
SNPOANNEIRQKLTSAVTKPPQGYPVQLSNDSTQKFIAKLESFAEGSRTAAEIK
ALSFTNSLFIQQVLVNLGSLYSGLQ"
complement(9770..12388)
/gene="CP0019"
complement(9770..12388)
/note="conserved hypothetical protein; identified by
Glimmer2; putative"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAF37915.1"
/db_xref="GI:7188958"
/translation="MSFESIONRLRTISGESTRIIKLDHKYSGDPRSPVPAINLEELN
SGIYALRHLNMAQSENVAALLNPNTIPTTSTWTDIKHRSPOASPRAPSTPT
DIVSAAALALVLDGGLAEVSAVTEIDGALSTIVRQLMASVGLTTLTAOEK
VVFSSYVPSEKLEHVKQKAAEIQAKAEIKAVLEAKGVSTEEIAILKEYPDYI
AADPFKEFIEPLTRAKVCAPIQEMNENAIQLPTTPTPDNNVNGMNTLSIKI
LOAIDDAIKQAPALGGQEEIITILQTLVPLVDKTFKAEFDLIYTATQLPNTASIKI
YLTDQTAIEYRGKITKYQNSIQNLSETKRVVENNRSMLETQLSHFQAOQCNFTWIS
QANALNTAITNKYISAVLTTSMEWYGLGCLSYMYERLAODEKAIQKSVNVEYLIPIHI
VGVSWNGWIAKMAAYQELAEYSLGTAVTSDQIKAYLOTRGNEFKATRHFFHIGD
OMYQFANETVEFNCGLTANGAIQPDLAGFIREAMTVGTVEADYVSNQAQRLNENFTA
ATAVHLOLOQIAELOKADDDPKGASTENRKFVAVAWITTSLSGLDALLSMILNSQ
LPKQEAFLKPLIEENFNLAANLSLLOITNEFSTTSVYSSLSLVQSTQGNLF

Query Match

92.5%; Score 1359.4; DB 1; Length 13548;

: Best Local Similarity 99.5%; Pred. No. 0;

Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;

Oy 1 atgggtaacccattggccaggtccctatagacgaacagacacacccctcccgagat 60

|||||

Db 7488 ATGGGTTAATCCATTATGGTCCAGGTCCTATAGACGAACAGACACCTCCCGAGAT 7547

Qy	61	ctttctgtc	taaggattg	gagcgag	tg	cagcaataa	gaagtg	tcggaag	ctc	aaagaata	120		
Db	7548	CTTTCTGCT	CAAGGAT	TGGAGCG	AGTGC	AGCAAA	TAAAGAT	TGCGGAAG	CTC	AAAGAA	7607		
Qy	121	gcaggtgc	ggaagc	taagcc	taagaat	ctaagacc	gatic	ctgtag	agc	atgcagc	atc	180	
Db	7608	GCAGGTGC	CGGAAGCT	TAAAGCT	TAAAGAT	TCTAAGAC	CCGAT	TCTGTAG	AGCGAT	TGGAGCAT	C	7667	
Qy	181	ttgcgtct	gcagtg	aaatgc	tctctcat	gagtc	gtgc	agataa	gctgg	ggtattg	ctt	tagt	240
Db	7668	TTGCGTCT	TGCAGT	GAATGCT	CTCATG	ATGCTG	GCAGATA	AAGCTG	GGTATTG	TCTTCT	AGT	7727	
Qy	241	aacagctg	ctctt	ctactag	atctc	gcagatc	ctgc	agacgt	ggact	caacg	acagc	gacgc	300
Db	7728	AACAGCTC	GCTCTT	CTACTAG	CAGAT	TCTCAG	ACGTG	GACTCA	ACGAC	GAGCG	AGCG	CAACCT	7787
Qy	301	acgcctct	cccccac	gctctg	atgatt	ataagact	caagcg	caaac	agctt	acgata	ctt	360	
Db	7788	ACGCTCT	CCCTCAC	CCGTTT	GATGATT	TAAAGACT	CAAGCG	CAACAG	CTTACG	GATACT	7847		
Qy	361	atctttac	ctoaac	tactag	ctg	acatac	agcg	ctctt	tg	tgagc	ctcc	aggtat	420
Db	7848	ATCTTTAC	CTCAACAT	CACTAG	CTGC	ATACAG	CTCAG	GCTGCT	TGTG	TAG	ACCTT	CC	7907
Qy	421	gtcactaa	tataa	agatac	agcg	gctact	gatg	agaaac	ccgaat	cgctg	cgag	gtg	480
Db	7908	GTCACTAA	TATAA	GGATAC	AGCGGCT	ACTGAT	GANGA	NAACCG	CACTCG	CTCGG	AGTGG	7967	
Qy	481	gaaactaa	gaatgc	gcg	atg	caataa	aaagt	gtgc	gcg	caaat	tcac	agaat	540
Db	7968	GAAACTAA	GAAATG	CCGAT	ACAGTT	TAAAGTT	TGCGCG	CGAAAT	TACAG	AAAT	TAG	CGAAAT	8027
Qy	541	gcttcgga	tataa	cccaag	catctt	gactctt	tag	taaac	agct	ctctt	cgc	acctta	600
Db	8028	GCTTCGGAT	AACCA	AGCGAT	CTTGAC	TCTTTAGG	TAAACT	GCATCT	CTCTCG	ACCTCT	TTA	8087	
Qy	601	cagactg	ctctt	ctcc	aatctg	tagca	acataa	caaac	agc	agctg	agctt	ctt	660
Db	8088	CAGGCTGC	TCTTCT	CCAACT	CTTAG	CAAA	CAAT	TACA	AAGC	AGCTG	AGCTT	CTT	8147
Qy	661	atcagaat	taac	cccg	atg	cccg	ggaaac	gc	ctg	caatt	gct	caatt	720
Db	8148	ATCGAGAT	TAACCC	AGTAG	TCC	CAGG	AAACG	CTG	CAAT	TGCT	CAAT	CTT	8207
Qy	721	cagacag	atg	ctac	agcg	acac	agatag	agaa	atg	gaaat	gcg	atg	780
Db	8208	CAGACAGAT	GTAC	ACG	CAC	CAGATAG	AGAA	GATG	AAAT	TGC	GATT	TAGG	8267
Qy	781	tttgagg	acag	aagc	tag	tgag	ctgt	agataa	atg	ctaaat	ctataa	ac	840
Db	8268	TTTTCAGG	ACAGAAC	GCTAGT	GGAGCT	GTAG	AAAA	TGTAAT	CTATA	TAC	AGTAA	GAAG	8327
Qy	841	aacatag	att	cagc	taag	cag	caat	cg	ctact	g	taag	acac	900
Db	8328	AACATAG	ATT	CAGTAA	ACG	CAATCG	CTACT	GTCTA	AGAC	CAAA	TAGCTG	AAAGCT	8387
Qy	901	aaaaagt	ctcc	gact	ctcc	aatctt	caag	agcg	gac	aaaaat	tggt	taatac	960
Db	8388	AAAAAGT	CTCCCG	ACTCT	CCAA	TCTT	CAAG	AGCG	GAA	CAAA	TGTAAT	CTAAC	8447
Qy	961	aaagatt	ctta	aaaa	atac	aaac	ctgc	agatg	gtt	ctg	atgt	tc	1020
Db	8448	AAAGATCT	TAAAA	ATATCA	AACT	CGAG	NTGG	TCTG	ATGTT	CCAA	TCC	AGGA	8507
Qy	1021	gttgagg	ctcc	caag	caac	aa	gaag	tag	tatt	gtg	tag	tatt	1080
Db	8508	GTTTGG	AGGCT	TCCA	AGCA	CAAG	CAAG	TAGT	ATT	TG	TAGT	ATT	8567
Qy	1081	gatgat	ctg	aaaa	atg	agacc	ctt	ccatt	ttg	atg	ctg	gg	1140
Db	8568	GATGAT	GCTG	AAAA	ATG	AGAC	CCGCTT	CCAT	ATTT	TGAT	GTCT	G	8627
Qy	1141	atgttca	atc	ac	gc	gaa	aat	ct	gatt	ctc	aa	gctg	1200

Db	8628	ATGTTCAATACGGAATTCCTGATTC	CAAGCTGCCAACAGGAGCTCGCAGCAAGCT	8687
QY	1201	agagcagcgaagccgctggagatgacagtgctgctgcagcgcctggcagatgctcagaaa	1260	
Db	8688	AGACACGGAAGCCGCTGGAGATGACAGTCTCTCTGCAGCGCTGGCAGATGCTCAGAAA	8747	
QY	1261	gcttagaagcggcctctagtgtaagcgtgggcaacaacaggcgcatactactcaatgctttggga	1320	
Db	8748	GCTTTAGAAGCGGCTTAGGTAAAGCTGGGCAACAACAGGCATACTCAATGCTTTAGGA	8807	
QY	1321	cagatccttctgctgctgtgttgtagcgacagag-tcctcccgctgcagcaggttctatg	1379	
Db	8808	CAGATCGCTTCTGCTGTGTGTGTCAGCGCAGAGTTCCTCCCGCTGCAGCAAGTTCATA	8867	
QY	1380	g	1380	
Db	8868	G	8868	
RESULT	6			
AP002547/c				
LOCUS		AP002547	300550 bp	DNA
DEFINITION		Chlamydomophila pneumoniae J138	genomic DNA, complete sequence,	BCT 08-DEC-2000
ACCESSION		AP002547	AB033786	AB033787
VERSION				
KEYWORDS		AP002547.2	GI:10176693	BA000008
SOURCE		Chlamydomophila pneumoniae J138 (strain:J138)	DNA.	
ORGANISM		Chlamydomophila pneumoniae J138		
REFERENCE		Bacteria; Chlamydiales; Chlamydiaceae; Chlamydomophila.		
AUTHORS		Shirai, M., Hirakawa, H., Kimoto, M., Tabuchi, M., Kishi, F., Ouchi, K., Shiba, T., Ishii, K., Hattori, M., Kuhara, S. and Nakazawa, T.		
TITLE		Comparison of whole genome sequences of Chlamydia pneumoniae J138 from Japan and CWL029 from USA		
JOURNAL		Nucleic Acids Res.	28 (12), 2311-2314	(2000)
MEDLINE				
REFERENCE				
AUTHORS		Shirai, M.		
TITLE		Direct Submission		
JOURNAL		Submitted (04-JUL-2000)	to the DDBJ/EMBL/GenBank databases.	
		Mutsunori Shirai, Yamaguchi University School of Medicine,		
		Department of Microbiology; 1-1-1 Minamikogushi, Ube, Yamaguchi		
		755-8505, Japan (E-mail:shirai@po.cc.yamaguchi-u.ac.jp,		
		Tel:81-836-22-2227, Fax:81-836-22-2415)		
COMMENT		On Sep 15, 2000 this sequence version replaced gi:6172298		
		gi:6172300 gi:6172396 gi:6172398 gi:8978889,		
		AB033786-AB033787, AB033816-AB033817. Submitted (25-Oct-1999).		
FEATURES		Location/Qualifiers		
source		1. 300550		
		/organism="Chlamydomophila pneumoniae J138"		
		/strain="J138"		
		/db_xref="taxon:138677"		
gene		/note="synonym:Chlamydia pneumoniae (strain J138)"		
		complement(21..794)		
		/gene="dapF"		
CDS		complement(21..794)		
		/gene="dapF"		
		/codon_start=1		
		/transl_table=11		
		/product="diaminopimelate epimerase"		
		/protein_id="BA098725.1"		
		/db_xref="GI:8978890"		
		/translation="MARYSPSTISKYFIYSAGNRFLLGELTPEVEDVRFICQETRVH		
		GFYLKPPSCADAQLIFNSDGSRTMGNGRLCAIAHLSQKSGKSDISVSDSLG		
		SYLFSWDLVDMTLADWRASVHLESRPDLPRKEVGIHTGVPHAVYILPEISLD		
		GIFGFLRYHOTFSPDGVNMFVQLIGHCOLRVTYERGVETGAACGTGALASALV		
		SNSYGVKESIQIHTWGGFLMTVSQNRGRVYLQGSVTRDL"		
gene		complement(763..1338)		
		/gene="clpP_1"		
CDS		complement(763..1338)		
		/gene="clpP_1"		

```

/codon_start=1
/transl_table=11
/product="CLP protease"
/protein_id="BAA98726.1"
/db_xref="GI:8978891"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98727.1"
/db_xref="GI:8978892"

```

gene

CDS

```

/translation="MLKVFEKPKPAIVEITKVAVVSVLLHKLLENASGKKGOSLAS
TAYLAADHLNAPSFISGERIIDELKSQSHLKIASENYSLSVQLAMGNLLTDKYC
EGSPFKRYSCCENVDATIEWECVETAKELFAADCACQVPHSGADNLAVMAILLTKV
QGPVSKLYKTVNELTEETVLLKAEMSSCVCLGPSLNSGGHETHGNVRLNVMSKLM
RCFPDVNPDTECPDYAEISLRAKEYPKVLIAGYSSYSRRLNFAVLKQIAEDCGSVL
WYMAHFAGLVAGGVFDEBNPIYADIVTTTTTHKLRGPRGLVLAETREYESTLKA
CPLMMGGPLPHVIAAKTVALKEAUSVDPKYAHQVNNARLAEERFISHLRLLTGGT
DNHMMVIDLGSIGTSGRTAEDILSVGVAVNRNLSLPSDAIGKWDTSIRLGTPTLTL
GMGIDEMEVEDADIIVKVLNRILRSCHVEGSSKKNKGELPEAIAQEARDRVNRLLRFP
LYPEIDLEALV"
3107..3778
/gene="CPj0522"
3107..3778
/gene="CPj0522"
/codon_start=1
/transl_table=11
/product="CT433 hypothetical protein"
/protein_id="BAA98728.1"
/db_xref="GI:8978893"
/translation="MTLVGLGNLQTKARKYQAHYLPILFFPYAKSTPONKRALQFLPQ
ATHVLTPSTHFLSRMTSLLSKATLTKTYLTCIGESTKERLLSFLGQVKYVYVATQ
ETAEIGFLLQALPSSARILYPHSLARPVIREFLYNRTFFSYPHYTVKPKLKNKI
LSKKYKIIFTSPSTVRAFAKIFPPPEPTYWCQGRMTLQEFQKTSQKQVSLLETGK
SRTPS"
3843..4175
/gene="CPj0523"
3843..4175
/gene="CPj0523"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98729.1"
/db_xref="GI:8978894"
/translation="MASSATPGFDGTAPSLFPATPRPNRYNFKLALFVTTAIALVWIAL
IATTIAGLCIHLCSFIFLTAIPLYFISRYICSHYARNVIALDVVDFHSLQDMRS
HSPIFSDF"
4223..5302
/gene="CPj0524"
4223..5302
/gene="CPj0524"
/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98730.1"
/db_xref="GI:8978895"
/translation="MSPGSRFTESSQVSVLSPVRDKETAPKQFTIAKISTAILIASL
ALGALVAGISITVILGNPFLALITITAFVSVTVFLVYHOMTSKVSNNQVLEONFK
PLGRAWQEKNDVCYNEMQFVNNHLPKFAIOTDASQFPQFTGLRLVIEKNQST
GLIFNPVGTNLNONTATNLTSTILYSTLKDKSVMDTCQREGGAKGDEPSPTEYVR
VKLPNEALDQFNLSAERKSLITFLGHVCGKSEELFNQEQYIQALLAYENCL
KALPESHAALVALPLFTSVYVPEEILPKEGTFYWDNQTAQFCKRALLDAIQNTALR
YPORESLIVLQDPNTIESQSRSE"
5642..6406
/gene="CPj0525"
5642..6406

```

gene

CDS

```

/codon_start=1
/transl_table=11
/product="hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"

```

gene

CDS

```

/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"

```

gene

CDS

```

/gene="CPj0525"
/codon_start=1
/transl_table=11
/product="CT398 hypothetical protein"
/protein_id="BAA98731.1"
/db_xref="GI:8978896"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98732.1"
/db_xref="GI:8978897"

```

gene

CDS

```

/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98733.1"
/db_xref="GI:8978898"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="serine hydroxymethyltransferase"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"

```

gene

CDS

```

/codon_start=1
/transl_table=11
/product="glutamate symport"
/protein_id="BAA98734.1"
/db_xref="GI:8978899"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"

```

gene

CDS

```

/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"
/translation="MADGEVHLRDIIEKELLEARRVEFPVTEKSASDAIKKLWYL
ELKDPGRFIVINSPGSGVAGFAVMDQIKMLTSPVTVVTVGLAASGVSLSCAAP
GRRATPSRIMIHPISGGITGTQATDLDIRAREILKTRAKIIDVYVEATNQPRDII
EKAIDRMWMTANAKDFGLDGLILFSPNDL"
complement(1358..2917)
/gene="glaA"
/complement(1358..2917)
/gene="glaA"
/codon_start=1
/transl_table=11
/product="ATPase"
/protein_id="BAA98735.1"
/db_xref="GI:8978900"

```

gene

CDS

TVVRRVSNAPQIFVKPTIASVWTHNGERIPKEALRELIRVGVFCGLGFPQGLNMLRE
EGIHGLKYLPLDRAHATKKELNYFCOAMROGGLLCTEKDSVKPLRLSGEVSLLP
IAKVEMLSVNQDDTLISLNNIEQIHKNRGN"

Query Match	92.5%	Score 1359.4	DB 1	Length 300550
Best Local Similarity	99.5%	Pred. No. 0		
Matches 1374	Conservative 0	Mismatches 6	Indels 1	Gaps 1
Qy 1	atgggttaacccattggtccaggctctatagacgaacacagacacacccctccgcagat	60		
Db 217598	ATGGTTAATCCCTATTGGTCCAGGCTCTATAGCAAAACAGAACGACACCTCCCGCAGAT	217539		
Qy 61	ctttctgctcaaggattgagcgcagtgacgaataagagtcggaagctcaagaata	120		
Db 217538	CTTTCTGCTCAAGGATTGAGCGGAGTGCAGCAATAAGAGTGCAGGAAGCTCAAGAATA	217479		
Qy 121	cgaggtcggaagctaaagcctaaagaatctaaagccgattctgttagacgcatggaacatc	180		
Db 217478	GCAGGTGGGAAGCTTAAGCCTTAAGAAATCTAAGACCGAATTCGTAGACGATGGAGCATC	217419		
Qy 181	ttgcgtctgcagtgatgctctcagctgagctgagcagataagctgggtattgctctagt	240		
Db 217418	TTGCGTTTGCAGTGAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT	217359		
Qy 241	aacagctogtcttactagcagatctgcagacgtggactcaacacagcagcagccact	300		
Db 217358	AACAGCTGCTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACACAGCAGCCGACCT	217299		
Qy 301	agcctcctccacccagctctgattattataagacatcgaagcgaacacagcttaacgatact	360		
Db 217298	AGCGCTCCTCCACCCAGCTTTGATGATTATAAGACTCAAGCGCAACACAGCTTACGATACT	217239		
Qy 361	atctttacctcaacatcagctagctgacatacagctgctttggtgagcctcagatgct	420		
Db 217238	ATCTTTACCTCAACATCACTACTACTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCT	217179		
Qy 421	gtcactaatataaaggatcacagcgcctactgatgaggaaacgcgaatcgctgcgagtg	480		
Db 217178	GTCACTAATATAAAGGATACAGCGGCTACTGATGAGGAACCGCAATCGCTCGGAGTGG	217119		
Qy 481	gaactaagaatgccagtgcaattaaagttggcgcgaataattacagaattagcgaatat	540		
Db 217118	GAACTAAGAATGCCAGTGCAGTTAAAGTTGGCGCGCAAAATACAGAATTAGCGAAATAT	217059		
Qy 541	gttcggataaacaagcagcttctgactctttgagtaactgaactctcttcgacctcta	600		
Db 217058	GCTTCGGATAACCAAGCGATTCTTGACCTTTAGTAACTGACCTCTCTCGACCTCTTA	216999		
Qy 601	cagactgctctctccaatctgtagcaacaataaacaagcagctgagcttcttaagag	660		
Db 216998	CAGGCTGCTCTTCTCCAACTCTGTAGCAACAATAACAAGCAGCTGAGCTTCTTAAGAG	216939		
Qy 661	atgcaagataacccagtagtccagggaaaacgctgcaattgctcaatcttttagttat	720		
Db 216938	ATGCAGATAACCCAGTAGTCCAGGGAAAACGCTGCAATTGCTCAATCTTTAGTTGAT	216879		
Qy 721	cagacagatgctacagcagacagatagagaagatggaatgcgattgggagatccat	780		
Db 216878	CAGACAGATGCTACAGCGACACAGATAGGAAGATGGAATGCCATAGGATGCATAT	216819		
Qy 781	tttcgaggacagaaacgctagtgagctgtagaaaaatgcataatcaataacagtagtaagc	840		
Db 216818	TTTGCAGGACAGACGCTAGTGGAGCTGTAGAAAATGCTAAATCTAATAACAGTATAAGC	216759		
Qy 841	acatatagcttaagcagacatcgctactgctgaagacacaataagctgaagctcag	900		
Db 216758	AACATAGATTACGTTAAAGCAGCAATCGCTACTGCTAGACACAAATAGCTGAAGCTCAG	216699		
Qy 901	aaaaagtcccgactctccaattctcaagagcggacaacaaatggttaatacagctcag	960		
Db 216698	AAAAAGTTCCCGACTCTCTCAATCTTCAAGAGCGGACAATAATGTTAATACAGGCTGAG	216639		
Qy 961	aaagatccttaaaaatatcaaacctgcagatggtctgtatgttctccaaatccaggaaactaca	1020		

Db 216638	AAAGATCTTAAAAATATCAAACTGCAGATGGTTCTGATGTTCCAAATCCAGGAACACTACA	216579		
Qy 1021	gttgagagctccaaagcaacgaagtagtagtattggttagtattcgtgttccatcgtctta	1080		
Db 216578	GTTGAGGCTCCAAGCAACAAGGAAGTAGTATTGTTAGTATTGTTTCCATGCTGTTA	216519		
Qy 1081	gatgatgctgaatatgagaccgctccatttgcatttgcattgctggtttcgcagatgattcac	1140		
Db 216518	GATGATGCTGAATGAGACCGCTTCCATTGTTGATGCTGGTTTCGTCAGATGATTCAC	216459		
Qy 1141	atgttcaatacggaaaaatccttgattctcaagctgcccacacagagctgcgcagcaagct	1200		
Db 216458	ATGTTCAATACGAAAAATCCTGATTCTCAAGCTGCCCAACAGGAGCTGCAGCACAGCT	216399		
Qy 1201	agagcagcgaagccgctggagatgacagatgctgctgagcagcgtggcagatgctcagaaa	1260		
Db 216398	AGAGCAGCGAAGCGCTGGAGATGACAGTCTGCTGCAGCGCTGGCAGATGCTCAGAAA	216339		
Qy 1261	gcttttagagcggtcttagttaaagctgggcaacacacagggcactactcaatgctttggga	1320		
Db 216338	GCTTTAGAAGCGGCTCTAGGTAAGCTGGGCAACAACAGGCGATCTCAATGCTTTAGGA	216279		
Qy 1321	cagatcgcttctgctgctgtgtgagcgcaggag-tctcccgctgagcaggtctctatg	1379		
Db 216278	CAGATCGCTTCTGCTGCTGTTGTGAGCGCAGGAGTTCCTCCGCTGCAGCAAGTTCTATA	216219		
Qy 1380	g 1380			
Db 216218	G 216218			
RESULT 7				
AX045135				
LOCUS	AX045135	1456 bp	DNA	PAT 24-NOV-2000
DEFINITION	Sequence 5 from Patent WO0066739.			
ACCESSION	AX045135			
VERSION	AX045135.1	GI:11343734		
KEYWORDS				
SOURCE	Chlamydia pneumoniae.			
ORGANISM	Chlamydia pneumoniae			
REFERENCE	1 (bases 1 to 1456)			
AUTHORS	Murdin,A.D., Omen,R.P., Wang,J. and Dunn,P.			
TITLE	i(chlamydia) antigens and corresponding dna fragments and uses thereof			
JOURNAL	Patent: WO 0066739-A 5 09-NOV-2000;			
FEATURES	Aventis Pasteur Limited (CA)			
source	Location/Qualifiers			
	1..1456			
	/organism="Chlamydia pneumoniae"			
	/db_xref="taxon:83558"			
	101..>1452			
	/note="unnamed protein product"			
	/codon_start=1			
	/transl_table=11			
	/protein_id="CAC17379.1"			
	/db_xref="GI:11343735"			
	/translation="MVPNIGPIDEITPPADLSAQGLEASANKSAEQAQTAGAE			
	AKPKESITDSVERMSILRSVAVNALMSLADKLGIASSNSSSTSRSDVDSTATATP			
	PPPTFDYKTAQATYDTIFTSTSLADIALVSLQDAVTNFKDTAATDEETAIAAEW			
	ETKNADVVKQAQITELAKYASDNQAILDLSLGLKTSFDLLQALQLSVANNKAAELL			
	KEMODNPVPGKTPAIQSLVDQDATOLEKDGNAITRDYAFYAGVAGAVAKSN			
	NSISNIISAKAATATATKTAEAKKEFPDPIQEAQEQMVIQAEKDLKNIKPAGSDV			
	PNQTTVGGSKQQGSSIGSIRVSMLLDDAENETASILMSGFRQMIHMFENTENPQAA			
	G" QOELAAQARAARAAAGDDSAALADAKPAALGKAGQQQGIILNALGQIASAAVWSA			
BASE COUNT	452 a 308 c 331 g 365 t			
ORIGIN				
Query Match	91.6%	Score 1347	DB 6	Length 1456
Best Local Similarity	99.6%	Pred. No. 0		

Matches 1350; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

```
Qy 1 atgggtaactctattggtccaggtcttatagacgaacacagacgcacacccctccgcagat 60
Db 101 ATGGTTAATCCTATTGGTCCAGGTCTTATAGAGAAACAGAACGCACACCTCCCGCAGAT 160
Qy 61 cttctgtcctaaggattgagggcagtgacgacaaataagagtgcggaagctccaaagaata 120
Db 161 CTTTCTGTCTAAGGATTGGAGCGAGTGCAGCAATAAAGAGTGGCGAAGCTCAAAAGAATA 220
Qy 121 cgagctgcggaagctaaagcctaaagaatactaaagaccgattctgttagagcgatggagc 180
Db 221 CGAGGTGCGGAAGCTTAAGCCTAAAGAACTTAAGACCGATTCTGTAGCGGATGGAGATC 280
Qy 181 ttgcgtctgcagtgaaatgctctcatgagctggtgagataagctgggtattgctcttagt 240
Db 281 TTGCCTTCTGCAGTCAATGCTCTCATGAGTCTGGCAGATAAGCTGGGTATTGCTTCTAGT 340
Qy 241 aacagctcgtctcttactagagatctgcagacgtggactcaaacacagcgacccagcct 300
Db 341 AACAGCTCGTCTTCTACTAGCAGATCTGCAGACGTGGACTCAACGACACGCGACCGCAGCT 400
Qy 301 acgctcctccaccacacgtctgatattataagactcaagcgcaaacagcttacgatact 360
Db 401 AGCCCTCTCCACCCACGCTTGTATGATTATAGACTCAAGCGCAACAGCTTACGATACT 460
Qy 361 atcttactcaacatcactagctgacatacagcgctgttggtagcctccagatgct 420
Db 461 ATCTTTTACTCAACTCACTAGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGGATGCT 520
Qy 421 gtcactaatataagatatagcggtcactgatgagggaaacgcgaatcgcgtcgcgagtg 480
Db 521 GTCACTAATAATAAGGATACAGCGGCTACTGATAGGAAACGCGAATCGCTGCGGAGTGG 580
Qy 481 gaaactaagaatgccgatgcaattaaacttgccgcgaattacagaattagcgaatat 540
Db 581 GAAACTAGAATGCCGATGCAGTTAAAGTTGGCGCGCAATATACAGAATTTAGCGAAATAT 640
Qy 541 gcttcggataacaaagcattcttgactctttagttaaactgaactccttcgaccttta 600
Db 641 GCTTCGGATAAACCAAGCATTCTTACCTTTAGTAAACTGACTTCTCTCGACCTCTTA 700
Qy 601 cagactcgtctctccaatctgttagcaacaaataaacaagcagctgagctctcttaagag 660
Db 701 CAGGCTGCTCTTCTCCAATCTGTAGCAACAATAACAAGCAGCTGAGCTTCTTTAAAGAG 760
Qy 661 atgcaagataaaccagtagtccaggggaaacgcctgcaattgctccaactcttagttgat 720
Db 761 ATGCAAGATACCCAGTAGTCCCAGGGAACGCCCTGCAATGTCTCAATCTTTAGTTGAT 820
Qy 721 cagacagatgctacagcgacacagatagagaaagatggaatgcgattggggatgcata 780
Db 821 CAGACAGATGCTTACAGCAGACACAGATAGAGAAAGATGGAATGCGATTAGGATGCATAT 880
Qy 781 ttgcagacagaacgcctagtggaactgtagaagaatgctaactcaataacagataagc 840
Db 881 TTTGAGACAGAACGCTAGTGGAGCTGTAGAAATGCTAAATCTAAATACAGATAGAGC 940
Qy 841 aacatagattcagctaaagcagcaatcgtactgctgaagacacaaatagctgaagctcag 900
Db 941 AACATAGATTACGTTAAAGCAGCAATCGCTACTGCTAGACACAATAAGCTGAAGCTCAG 1000
Qy 901 aaaaagtcccgactcctccaatcttcaagaagcgggaacaaatggtaatacaggtgag 960
Db 1001 AAAAAAGTTCCCGACTCTCCAATTTCTCAAGAAAGCGGAACAAATGCTAATACAGGCTGAG 1060
Qy 961 aaagatcttaaaaatatcaacctcagatggttctgattgttccaaatccagggaactaca 1020
Db 1061 AAAGATCTTAAAAATATCAAACTCGATGCTTCTGATGTCTCAATTTCCAAATCCAGGAATACA 1120
Qy 1021 gttggaggtcccaagcaacaaggaatgattggttagtattcgtgttccatcgtgtta 1080
Db 1121 GTTGAGGCTCCAAACACACAGGAAGTAGTATTGGTAGTATTCGTGTTCATGCTGTGTA 1180
```

```
Qy 1081 gatgatgctgaaaaatgagaccgcttccattttgatgtctgggttttcgcatgattcac 1140
Db 1181 GATGATGCTGAAAAATGAGACCGCTCCATTTCATTTGATGCTGGGTTTCGTGATGATTCAC 1240
Qy 1141 atgttcaatacggaaaaatccttgattctcaagctgccccacaggagctgcgacgacaagct 1200
Db 1241 ATGTTCAATAGGAAAAATCCTGATTCTCAAGCTGCCAACAGAGAGCTGCCAGCACAAGCT 1300
Qy 1201 acagcagcaaaacgcgctggagatgacagtctgctcagcgcgtgagatgctcagaaa 1260
Db 1301 AGAGCAGCGAAAGCGCTGGAGATGACAGTCTGCTGCAGCGCTGGCAGATGCTCAGAAA 1360
Qy 1261 gcttttagaagcggtcctaggtaaaagctgggcaacacaggggcatactcaatgctttggga 1320
Db 1361 GCTTTTGAAGCGGCTCTAGGTAAGCTGGCAACAACAGGCGATACTCAATGCTTTAGGA 1420
Qy 1321 cagatcgcttctgctgctgtgtgtagcgacgagagt 1355
Db 1421 CAGATCGCTTCTGCTGCTGTTGTGAGCGCAGGAGT 1455
```

RESULT 8

AX045133 1852 bp DNA PAT 24-NOV-2000

DEFINITION Sequence 3 from Patent WO0066739.

ACCESSION AX045133

VERSION AX045133.1 GI:11343732

KEYWORDS

SOURCE Chlamydomophila pneumoniae.

ORGANISM Chlamydomophila pneumoniae

REFERENCE 1 (bases 1 to 1852)

AUTHORS Murdin, A.D., Oomen, R.P., Wang, J. and Dunn, P.

TITLE i(chlamydia) antigens and corresponding dna fragments and uses

JOURNAL thereof

PATENT: WO 0066739-A 3 09-NOV-2000;

Aventis Pasteur Limited (CA)

Location/Qualifiers

FEATURES

source

1. 1852

/organism="Chlamydomophila pneumoniae"

/db_xref="taxon:83558"

1. 1752

/note="unnamed protein product"

/codon_start=1

/transl_table=11

/protein_id="CAC17378.1"

/db_xref="GI:11343732"

/translation="MSLADKLGIASSNSSSSRSADVDTTATATPTPPPTFDDYKT

QAQAYDTIFTSTSLADIQAALVSLQDAVTNFKDTAATDEETAABWETKNADAVKY

GAQITELAKYASDNQAIIDSLGKLTSLDQALLQSVANNKAAELIKEMQDNPVP

GKPTAIQAQSLVDQTDATATQTEKQGNARIDYAFAGQNASGAVENAKSNNSIDSAK

AAIATANTQIAEAOAKFPDPSILQEAQOMVIOAEKDLNLIKPADGSDVPNGPTTVGGS

KOQSSISGIRVSMLLDDAENETASILMSGFROMIHMENTPDSQAQOPELAQAARA

AKAGDSRAAALADAKALEAALGKAGQOQILNALGQIASAAVYSGVPPAAASST

GSSVYKLYKTSKSGSDIKTQISAGYISDINAYGRNARDRVINNNPALTALR

SVPARTAEARPEKTDQALARVTSNRSRTLGVDYVSQVSAQSVNQITQSNPQANNEI

RQKLTSATVTPPOFGYPYVQLSNDSTQKFTAKLESLEFAEGSRTAAEIKALSFETNSLF

IQOVLVNIQSLYGLQ"

BASE COUNT 578 a 409 c 407 g 458 t

ORIGIN

Query Match 78.6%; Score 1155.4; DB 6; Length 1852;

Best Local Similarity 99.4%; Pred. No. 9.3e-271;

Matches 1170; Conservative 0; Mismatches 5; Indels 1; Gaps 1;

Qy 205 atgagctcggcagataaagctgggtattgcttcttagtaaacagctcgtcttctactagcaga 264

Db 1 ATGAGCTCGGAGATAAGCTGGGTATTGCTTCTAGTAACAGCTCGCTTCTACTAGCAGA 60

Qy 265 tctgcagagctggactcaacgacagcgaccgcacctaagctcttccaccacagctcgtgat 324

Db 1 tctgcagagctggactcaacgacagcgaccgcacctaagctcttccaccacagctcgtgat 324

Db 61 TCTGCAGACGTGGACTCAACACACAGCGACCGACCTAGCTCTCTCCACCCACCGTTTGAT 120
Qy 325 gattataagactcaagcgcaaaagcttaacgataactatcttttaactcaactcaactagct 384
Db 121 GATTATAAGACTCAAGCGCAACAGCTTACGATACCTATCTTTACTCAACATCACTAGCT 180
Qy 385 gacatacagcgctgtcttggtgagctccagagctgtgctcactaataataaagagatacagcg 444
Db 181 GACATACAGGCTGTCTTGGTGAGCTCCAGGATGCTGTCTACTAATATAAAGGATACAGCG 240
Qy 445 gctactgatgaggaacccgaatcgctgagtgaggagtggaacactaagaatgcccagtcgaatt 504
Db 241 GCTACTGATGAGGAACCGCAATCGCTCGGAGTGGGAACCTAAGAAATGCCGATCGAGTT 300
Qy 505 aaagttgagcgcaaaattacagaattagcgaataatgcttcggataacccaagcattctt 564
Db 301 AAAGTTGCGCGCAAAATACAGAATTAGCGAAATATGCTTCGGATAACCAAGCGATTCTT 360
Qy 565 gactcttttagttaaactgactctcttcgacctcttaacagctgctcttctccaactctgta 624
Db 361 GACTCTTTAGTAAACTGACTCTCTTCGACCTCTTACAGGCTGCTCTTCTCCAATCTGTA 420
Qy 625 gcaacaataacaaagcagctgagctctttaaagagatgcaagataacccagtagtccca 684
Db 421 GCAACAATAACAAAGCAGCTGAGCTTCTTAAAGAGATGCAAGATAACCCAGTAGTCCCA 480
Qy 685 gggaaaaacgctgcaattgctcaatcttttagttgatacagacagatgctacagcgacacag 744
Db 481 GGGAAAACGCTGCAATTGCTCAATCTTTAGTTGATCAGACAGATGCTTACAGCGACACAG 540
Qy 745 atagagaagatggaatcgattgggagtgatgcatattttgaggacagaaacgctagtgga 804
Db 541 ATAGAGAAGATGGAATCGGATTTAGGATGCAATTTTGGCAGACAAAGCTAGTGA 600
Qy 805 gcttagaaaaatgctaaatcttaataacagtagtaagcaacatagattcagctaaagcagca 864
Db 601 GCTGTAGAAAATGCTAAATCTAATAACAGTATAAAGCAACATAGATTACGCTAAAGCAGCA 660
Qy 865 atcgctactgtaagacaaatagctgaactcagaactcagaagaagttcccgactctccaatt 924
Db 661 ATCGCTACTGCTTAAGACAAATAGCTGAAGCTCAGAAAAGTTCCCGGACTCTCCAATT 720
Qy 925 cttaagaagcggaacaaatgtaatacagcgctgagaagatctttaaaaaatatacaaacct 984
Db 721 CTTCAAGAGCGGAACAATGTAATACAGGCTGAGAAAGATCTTTAAAAATATCAAACT 780
Qy 985 gcagatggtctgtatgttccaaatccaggaactacagttgaggtcccaagcaacaagga 1044
Db 781 GCAGATGTTCTGTATGTTCCAAATCCAGGAACACTACAGTTGGAGGCTCCAAGCAACAAGGA 840
Qy 1045 agtagtattgtaattatgctgtttccatgctgttagatgactgaaatgagaccgt 1104
Db 841 AGTAGTATTGGTAGTATTGCTGTTTCCATGCTGTAGATGATGCTGAAAATGAGACCGCT 900
Qy 1105 tccattttgattctgggtttgctcagatgattccatgttcaatcacggaaaatcctgat 1164
Db 901 TCCATTTGATGCTGGGTTTCGTTCAGATGATTCATGTTCAATACGGAATCTCTGAT 960
Qy 1165 tctcaagctgcccacagagctgcagcaacagtagacgagcgaagccgctggagat 1224
Db 961 TCTCAAGCTGCGCAACAGAGGCTCGCAGCACCAAGCTAGACGAGCGAAAGCCGCTGGAGAT 1020
Qy 1225 gacagtgtctgcagcgctgcaatgctcagaagctttagaagcggtctaggtaaa 1284
Db 1021 GACAGTGTCTGTCAGCGCTGCAGATGCTCAGAAAGCTTTTGAAGCGGCTCTAGGTTAA 1080
Qy 1285 gctgggcaacaacagggcatactcaatgctttgggacagatcgcttctgctgtgttg 1344
Db 1081 GCTGGGCAACAACAGGGCATACTCAATGCTTTAGGACAGATCGCTTCTGCTGCTGTGTG 1140
Qy 1345 agcgagggag-tctctcccgctgcagcaagttcttatgg 1380
Db 1141 AGCGCAGGAGTTCTCTCCCGCTGCAGCAAGTTCTATAG 1177

RESULT 9
ARI144067 150 bp DNA PAT 08-AUG-2001
LOCUS Sequence 8 from patent US 6210876.
DEFINITION ARI144067
ACCESSION ARI144067
VERSION ARI144067.1 GI:15105934
KEYWORDS
SOURCE Unknown.
ORGANISM Unclassified.
REFERENCE 1 (bases 1 to 150)
AUTHORS Cerney, M.B.
TITLE Nucleic acid primers and probes for detecting Chlamydia pneumoniae
JOURNAL Patent: US 6210876-A 8 03-APR-2001;
FEATURES Location/Qualifiers
source 1..150
BASE COUNT 43 a 34 c 40 g 33 t
ORIGIN

Query Match 10.0%; Score 146.8; DB 6; Length 150;
Best Local Similarity 98.7%; Pred. No. 4.2e-25;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Qy 366 tacctcaactcaactagctgacatacaggtgcttttgtagcctccaggatgctgtcac 425
Db 1 TACTCTAACATFCATAGCTGACATACAGGCTGCTTTGGTGAGCTCCAGGATGCTGTAC 60
Qy 426 taataaagatcacagcgctactgatgaggaacccgaatcgctcgagtgaggaaac 485
Db 61 TAATATAAGATACAGCGGCTACTGATGAGAAACCCCAATCGCTCGGTGTGGAAAC 120
Qy 486 taagaatgcgagtgcaattaaagtggcgc 515
Db 121 TAAGAATGCCGATGCAGTTAAAGTTGGCGC 150

RESULT 10
AX105043 816 bp DNA PAT 30-APR-2001
LOCUS Sequence 19 from Patent WO0125445.
DEFINITION AX105043
ACCESSION AX105043
VERSION AX105043.1 GI:13921205
KEYWORDS
SOURCE Agkistrodon contortrix.
ORGANISM Agkistrodon contortrix
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroidea;
Viperidae; Crotalinae; Agkistrodon.
REFERENCE 1 (bases 1 to 816)
AUTHORS Boone, T.C., Li, H. and Mann, M.B.
TITLE Fibrinolytically active polypeptide
JOURNAL Patent: WO 0125445-A 19 12-APR-2001;
Angen Inc. (US)
FEATURES Location/Qualifiers
source 1..816
/organism="Agkistrodon contortrix"
/db_xref="taxon:8720"
/note="Fragment of fibrinolyase of Agkistrodon contortrix"

BASE COUNT 220 a 219 c 168 g 209 t
ORIGIN

Query Match 4.4%; Score 65; DB 6; Length 816;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Qy 1406 tagaacaacaaactcaatcagaagaggtctgaatagcgccgtcgaccatcatcatc 1465
Db 646 TAGAACAAAACCTCATCTCAGNAGAGGATCTGAATAGGCCGCTCGACCATCATCATC 705

QY 1466 atcat 1470
|||||
Db 706 ATCAT 710

RESULT 11
AX092220
LOCUS AX092220 1200 bp DNA PAT 21-MAR-2001
DEFINITION Sequence 36 from Patent WO0116321.
ACCESSION AX092220
VERSION AX092220.1 GI:13444420
KEYWORDS human.
SOURCE Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1200)
AUTHORS Tandon, N., Sun, B., Nakamura, T. and Yamamoto, N.
TITLE Platelet membrane glycoprotein vi (gpvi) dna and protein sequences,
and uses thereof
JOURNAL Patent: WO 0116321-A 36 08-MAR-2001;
OTSUKA PHARMACEUTICAL CO., LTD. (JP)
FEATURES
source Location/Qualifiers
1..1200
/organism="Homo sapiens"
/db_xref="taxon:9606"
41..1150
/note="unnamed protein product"
/codon_start=1
/protein_id="CAC34914.1"
/db_xref="GI:13444421"
/translation="MRPFSIFAVLFAASSALAAPVNTTDETAQIPAEAVIGYSDI
EGEDVAVLPFSNTNNGLLFINTTIAAAKEGVSLKREAEFTWPSRQSGPL
PKSLQALPSSLVLEKPVTLRCQPGVDLYLEKLSRRYQDAVLFIAMKRSLSA
GRYCSYONGSLWLSLPSDLELVATGVFAKPSLSAOPGPVSSGGDVTLOQTRGFD
QFALYKGGDPAPYKNPWPYRASPILITVTAHSGTVRCYSFSSRDPLWAPSDPLE
LVVGTSTPSRLTEPPSSVAEFSAEATLTVSTNKFVFTTSRSTTSPPKESDSP
AGPARYITKNGLEQKLISEEDLSNAVDDHHHHH"
BASE COUNT 296 a 361 c 285 g 258 t
ORIGIN

Query Match 4.4%; Score 65; DB 6; Length 1200;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 tagaacaacatctctcagaagaggtctgaatagcgccgtcgaccatcatcatc 1465
|||||
Db 1083 TAGACAAAACATCTCTCAGAGAGGATCTGAATAGCGCGTCGACCATCATCATC 1142

QY 1466 atcat 1470
|||||
Db 1143 ATCAT 1147

RESULT 12
AX105044
LOCUS AX105044 1373 bp DNA PAT 30-APR-2001
DEFINITION Sequence 20 from Patent WO0125445.
ACCESSION AX105044
VERSION AX105044.1 GI:13921206
KEYWORDS Agkistrodon contortrix.
SOURCE Agkistrodon contortrix
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroides;
Viperidae; Crotalinae; Agkistrodon.
REFERENCE 1 (bases 1 to 1373)
AUTHORS Boone, T.C., Li, H. and Mann, M.B.
TITLE Fibrinolytically active polypeptide
JOURNAL Patent: WO 0125445-A 20 12-APR-2001;
Amgen Inc. (US)

Query Match 4.4%; Score 65; DB 6; Length 1200;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 tagaacaacatctctcagaagaggtctgaatagcgccgtcgaccatcatcatc 1465
|||||
Db 1083 TAGACAAAACATCTCTCAGAGAGGATCTGAATAGCGCGTCGACCATCATCATC 1142

QY 1466 atcat 1470
|||||
Db 1143 ATCAT 1147

FEATURES
source Location/Qualifiers
1..1373
/organism="Agkistrodon contortrix"
/db_xref="taxon:8720"
/note="Fragment of fibrolase of Agkistrodon contortrix"
BASE COUNT 395 a 329 c 291 g 358 t
ORIGIN

Query Match 4.4%; Score 65; DB 6; Length 1373;
Best Local Similarity 100.0%; Pred. No. 3.2e-05;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 tagaacaacatctctcagaagaggtctgaatagcgccgtcgaccatcatcatc 1465
|||||
Db 1203 TAGACAAAACATCTCTCAGAGAGGATCTGAATAGCGCGTCGACCATCATCATC 1262

QY 1466 atcat 1470
|||||
Db 1263 ATCAT 1267

RESULT 13
AX105036
LOCUS AX105036 1620 bp DNA PAT 30-APR-2001
DEFINITION Sequence 12 from Patent WO0125445.
ACCESSION AX105036
VERSION AX105036.1 GI:13921202
KEYWORDS Agkistrodon contortrix.
SOURCE Agkistrodon contortrix
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Lepidosauria; Squamata; Scleroglossa; Serpentes; Colubroides;
Viperidae; Crotalinae; Agkistrodon.
REFERENCE 1 (bases 1 to 1620)
AUTHORS Boone, T.C., Li, H. and Mann, M.B.
TITLE Fibrinolytically active polypeptide
JOURNAL Patent: WO 0125445-A 12 12-APR-2001;
Amgen Inc. (US)

FEATURES
source Location/Qualifiers
1..1620
/organism="Agkistrodon contortrix"
/db_xref="taxon:8720"
/note="Complementary (sense) strand of antisense strand
(See SEQ ID NO:13)
Coding sequence of native pro-fibrolase of Agkistrodon
contortrix"
BASE COUNT 465 a 379 c 337 g 439 t
ORIGIN

Query Match 4.4%; Score 65; DB 6; Length 1620;
Best Local Similarity 100.0%; Pred. No. 3.1e-05;
Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1406 tagaacaacatctctcagaagaggtctgaatagcgccgtcgaccatcatcatc 1465
|||||
Db 1412 TAGACAAAACATCTCTCAGAGAGGATCTGAATAGCGCGTCGACCATCATCATC 1471

QY 1466 atcat 1470
|||||
Db 1472 ATCAT 1476

RESULT 14
AX105037/c
LOCUS AX105037 1620 bp DNA PAT 30-APR-2001
DEFINITION Sequence 13 from Patent WO0125445.
ACCESSION AX105037
VERSION AX105037.1 GI:13921203
KEYWORDS Agkistrodon contortrix.
SOURCE Agkistrodon contortrix.

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 19:50:20 ; Search time 685.35 seconds
(without alignments)
1838.868 Million cell updates/sec

Title: US-09-391-606-3
Perfect score: 1470
Sequence: 1 atgttaactctattgtcc.....accatcatcatcatcatcat 1470

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 930621 seqs, 428662619 residues

Total number of hits satisfying chosen parameters: 1861242

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : N_Geneseq_1101.*
1: /SIDS2/gcgdata/geneseq/geneseq/NA1980.DAT.*
2: /SIDS2/gcgdata/geneseq/geneseq/NA1981.DAT.*
3: /SIDS2/gcgdata/geneseq/geneseq/NA1982.DAT.*
4: /SIDS2/gcgdata/geneseq/geneseq/NA1983.DAT.*
5: /SIDS2/gcgdata/geneseq/geneseq/NA1984.DAT.*
6: /SIDS2/gcgdata/geneseq/geneseq/NA1985.DAT.*
7: /SIDS2/gcgdata/geneseq/geneseq/NA1986.DAT.*
8: /SIDS2/gcgdata/geneseq/geneseq/NA1987.DAT.*
9: /SIDS2/gcgdata/geneseq/geneseq/NA1988.DAT.*
10: /SIDS2/gcgdata/geneseq/geneseq/NA1989.DAT.*
11: /SIDS2/gcgdata/geneseq/geneseq/NA1990.DAT.*
12: /SIDS2/gcgdata/geneseq/geneseq/NA1991.DAT.*
13: /SIDS2/gcgdata/geneseq/geneseq/NA1992.DAT.*
14: /SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT.*
15: /SIDS2/gcgdata/geneseq/geneseq/NA1994.DAT.*
16: /SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT.*
17: /SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT.*
18: /SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT.*
19: /SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT.*
20: /SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT.*
21: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT.*
22: /SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query	Score	Match	Length	DB	ID	Description
1	1460.4	99.3	2238	21	AA02066		C. pneumoniae 76 k
2	1359.4	92.5	2156	21	AA02063		Chlamydia pneumoniae
3	1348.4	91.7	1230025	20	AA02063		Nucleotide sequenc
4	1347	91.6	1456	21	AA02065		3'-truncated Chlam
5	1155.4	78.6	1852	21	AA02064		5'-truncated Chlam
6	146.8	10.0	150	19	AAV16207		Part of the gene e
7	83	5.6	936	22	AAF58252		Oligonucleotide D1
8	83	5.6	936	22	AAF58254		Oligonucleotide D1
9	83	5.6	936	22	AAF58257		Oligonucleotide D1
10	83	5.6	936	22	AAF58259		Oligonucleotide D2
11	83	5.6	936	22	AAF58262		Oligonucleotide D2

12	83	5.6	938	22	AAF58255		Oligonucleotide D1
13	80.6	5.5	936	22	AAF58252		Oligonucleotide D1
14	80.6	5.5	936	22	AAF58254		Oligonucleotide D1
15	80.6	5.5	936	22	AAF58257		Oligonucleotide D1
16	80.6	5.5	936	22	AAF58259		Oligonucleotide D2
17	80.6	5.5	936	22	AAF58262		Oligonucleotide D2
18	80.6	5.5	938	22	AAF58255		Oligonucleotide D1
19	79.4	5.4	2679	21	AA02066		Soluble neuropilin
20	75.6	5.1	2670	21	AA02067		Soluble neuropilin
21	74.8	5.1	2230	22	AA02067		Human cervical can
22	70	4.8	1550	21	AA02067		Chlamydia pneumoniae
23	65	4.4	816	22	AA02067		Agkistrodon contor
24	65	4.4	1200	22	AA02067		Agkistrodon contor
25	65	4.4	1373	22	AA02067		Agkistrodon contor
26	65	4.4	1620	22	AA02067		Agkistrodon contor
27	65	4.4	1620	22	AA02067		Agkistrodon contor
28	63	4.3	1909	21	AA02067		Rhesus monkey meia
29	63	4.3	4357	22	AA02067		Expression vector
30	63	4.3	7108	20	AA02067		EP-892047 Seq ID 3
31	63	4.3	7475	20	AA02067		EP-892047 Seq ID 3
32	63	4.3	8192	20	AA02067		EP-892047 Seq ID 3
33	56.4	3.8	72	20	AA02067		PI35-P13 construct
34	53.6	3.6	1537	21	AA02067		C. trachomatis LCV
35	53.6	3.6	1537	22	AA02067		Chlamydia trachoma
36	50.6	3.4	896	15	AA02067		PspA protective an
37	50.6	3.4	946	20	AA02067		Streptococcus pneu
38	50.6	3.4	957	18	AA02067		Streptococcus pneu
39	50.6	3.4	1990	18	AA02067		Streptococcus pneu
40	50.6	3.4	2085	15	AA02067		Pneumococcal surfa
41	50.6	3.4	2085	17	AA02067		DNA encoding pneum
42	50.6	3.4	2085	17	AA02067		Pneumococcal surfa
43	50.6	3.4	2085	19	AA02067		Pneumococcal surfa
44	50.6	3.4	2085	19	AA02067		Streptococcus pneu
45	50.6	3.4	2085	20	AA02067		Streptococcus pneu

ALIGNMENTS

RESULT	1
ID	AA02066
XX	AA02066 standard; DNA; 2238 BP.
AC	AA02066;
DT	26-MAR-2001 (first entry)
XX	C. pneumoniae 76 kDa protein truncation mutant fusion gene.
DE	76 kDa protein; bactericidal; diagnosis; prevention;
KW	pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
KW	acute respiratory disease; cough; sore throat; hoarseness; fever;
KW	vaccine; immunisation; treatment; truncation mutant; fusion gene;
KW	ds.
OS	Chlamydia pneumoniae.
OS	Synthetic.
FH	Key
FT	misc_feature
FT	Location/Qualifiers
FT	/tag= a
FT	/note= "This part of the sequence is unrelated to
FT	C. pneumoniae 76 kDa gene"
FT	766..2238
FT	/tag= b
FT	/product= "Truncated Chlamydia pneumoniae
FT	76kDa protein"
FT	2122..2238
FT	/tag= c
FT	/note= "This part of the sequence is unrelated to
FT	C. pneumoniae 76 kDa gene"
XX	
PN	WO2000065739-A2.

XX	Chlamydia pneumoniae 76 kDa full-length protein encoding DNA.
DE	
XX	76 kDa protein; bactericidal; diagnosis; prevention; treatment;
KW	pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
KW	acute respiratory disease; cough; sore throat; hoarseness; fever;
KW	vaccine; immunisation; ds.
XX	
OS	Chlamydia pneumoniae.
XX	
Key	Location/Qualifiers
FT	101..2056
CDS	/*tag= a
FT	/product= "Chlamydia pneumoniae 76kDa protein"
FT	
XX	
PN	W0200066739-A2.
XX	
PD	09-NOV-2000.
XX	
PF	03-MAY-2000; 2000WO-CA00511.
XX	
PR	03-MAY-1999; 99US-0132270.
PR	30-JUN-1999; 99US-0141276.
XX	
PA	(AVET) AVENTIS PASTEUR LTD.
XX	
PI	Murdin AD, Oomen RP, Wang J, Dunn P;
XX	
DR	WPI; 2000-687542/67.
DR	P-PSDB; AAY71954.
XX	
PT	Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
PT	useful for vaccinating against Chlamydia infections -
XX	
PS	Claim 2a; Fig 1; 112pp; English.
XX	
CC	The present sequence is a DNA coding for Chlamydia pneumoniae
CC	76 kDa protein. C. pneumoniae 76 kDa protein
CC	is used in the diagnosis, prevention and treatment
CC	of C. pneumoniae infections. (e.g. pneumonia, upper respiratory
CC	tract disease, bronchitis, sinusitis and acute respiratory
CC	disease such as cough; sore throat, hoarseness, fever; and
CC	abnormal chest sounds on auscultation). C. pneumoniae sequence
CC	is also used as vaccines for immunising humans against diseases
CC	caused by C. pneumoniae.
XX	
SQ	Sequence 2156 BP; 672 A; 461 C; 471 G; 552 T; 0 other;
Query Match 92.5%; Score 1359.4; DB 21; Length 2156;	
Best Local Similarity 99.5%; Pred. No. 0;	
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps	
Qy	1 atggttaattcctattggtccaggtcctatagacgaacacgaacacacccgcagat 60
Db	101 atggttaattcctattggtccaggtcctatagacgaacacgaacacccgcagat 160
Qy	61 ctcttgctcaaggattggaggcgagtcgagcaataaagagtcggaagctcaagaata 120
Db	161 ctcttgctcaaggattggaggcgagtcgagcaataaagagtcggaagctcaagaata 220
Qy	121 gcagggtcggaagctaagcctaagaattagaaccgattctgtagagcgatggagcattc 180
Db	221 gcagggtcggaagctaagcctaagaattagaaccgattctgtagagcgatggagcattc 280
Qy	181 ttgcgttctcagtggaatgctctcattgcttgcgcagataaagctgggtattgtcttctagt 240
Db	281 ttgcgttctcagtggaatgctctcattgcttgcgcagataaagctgggtattgtcttctagt 340
Qy	241 aacagctcgtcttctactagcagatctgcagacgttggtactcaacgacgaccgcacct 300
Db	341 aacagctcgtcttctactagcagatctgcagacgttggtactcaacgacgaccgcacct 400

Db 1481 g. 1481

RESULT 3
ID AAX91990/c
AXX91990 standard; DNA; 1230025 BP.

XX AC AAX91990;
XX DT 13-SEP-1999 (first entry)

XX Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
DE
XX
XX Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; ss.
XX Chlamydia pneumoniae.
OS
XX
XX W09927105-A2.
PN
XX
XX 03-JUN-1999.
PD
XX
XX 20-NOV-1998; 98WO-IB01890.
PF
XX
XX 04-NOV-1998; 98US-0107078.
PR
XX 21-NOV-1997; 97FR-0014673.
PR
XX (GEST) GENSET.
XX
XX
XX Griffais R;
PI
XX
XX WPI; 1999-357842/30.
DR
XX
XX Genome sequence of Chlamydia pneumoniae
PS
XX Claim 1; Page 291-611; 1912pp; English.

QY 241 aacagctcgtcttactagcagatctgagacgtgagctcaacgacgagccgcaact 300
Db 828358 AACAGCTCGTCTTACTAGCAGATCTGAGACGTGGACTCAACGACGCGACACCT 828299
QY 301 acgctctctccaccocagctgtgattataagactcaagcgcacaaacagcttacataact 360
Db 828298 ACGCCCTCCACCCACGTTTATGATGTTTAAAGACTCAAGCGCAACACAGCTTACGATACT 828239
QY 361 atctttacctcaacatcactagctacacagctgcttctggtgagctccagatgct 420
Db 828238 ATCTTTACCTCAACACTAGCTAGCTGACATACAGGCTGCTTTGGTAGCCTCCAGGATGCT 828179
QY 421 gtccactaataaagatacagcggctactgatgaggaacccgcaatcgctgcggagtgg 480
Db 828178 GTCACATAATAAAGGATACAGCGGTACTGATGAGAAACCGCAATCGCTGCGGASTG 828119
QY 481 gaaactaagaatccgagtgcaattaaagtggcgcaaatcacgaattacagaataat 540
Db 828118 GAAACTAAGAATCCGATGCGATTAAAGTTGGCGCGCAATTAACGAATTAGGCAATAT 828059
QY 541 gcttcgataaccaagcgattctgactcttttagttaaacgactctcttcgaccttta 600
Db 828058 GCTTCGGATAACCAAGGATCTTACTCTTTAGGTAAGTACTCTTCGACCTCTTA 827999
QY 601 cagactgctcttctccaatctgtagcaacaataaacaagcagctgagcttcttaagag 660
Db 827998 CAGGCTGCTCTTCTCCAAATCTGTAGCAACAATAAACAAAGCAGCTGAGTCTTAAAGAG 827939
QY 661 atcgaagataaccagtagtcccgaggaacccgctcaattctgctcaattcttagttgat 720
Db 827938 ATGCAAGATAACCCAGTAGTCCCGAGGAAACGCTTCAATTTGCTCAATCTTTAGTTGAT 827879
QY 721 cagacagatgctacagcagacacagatagagaagatggaatgcgattggggatgcata 780
Db 827878 CAGACAGATGCTACAGCGACACAGATAGAGAAAGATGGAATGCGATTAGGATGCATAT 827819
QY 781 ttgcaggaacagacgctagtgagctgtgagaaatctaaatctaaatctaaatctaaatc 840
Db 827818 TTTGAGGACAGACGCTAGTGGAGCTGTAGAAATGCTAAATCTAATACAGTATTAAGC 827759
QY 841 aactagattcagctaaagcagcaatcgctactgctaaagacacaaatagctgaagctcag 900
Db 827758 AACATAGATTGAGTAAAGCAGCAATCGCTACTGCTAAGACACAAATAGCTGAAGCTCAG 827699
QY 901 aaaagtctcccgactctccaatcttcaagaagcgaacaaatggtaatacagctcag 960
Db 827698 AAAAGTTTCCCGACTCTCCAATTTCTCAAGAAGCGGAACAAATGGTAATACAGGCTGAG 827639
QY 961 aaagatcttaaaatatacaaacctgagatggttctgagttccaaatccaggaactaca 1020
Db 827638 AAAGATCTTAAATATCAAACTGACATGTTCTGATGTTCCAAATCCAGGAATACA 827579
QY 1021 gttgaggtctcaagcaacaaaggaagtagtattggttagtattctggttccatgctgta 1080
Db 827578 GTTGGAGGCTCCACAGCAACAGGAAGTAGTATTGGTAGTATTTCGTTTCCATGCTGTTA 827519
QY 1081 gatgatgtgaaatgagacgcttccattttgagttgctggttttcgcatgattcac 1140
Db 827518 GATGATGCTGAAATATGACACCGCTTCCATTTTGTGTTCTGGGTTTCGTCAGATGATTCAC 827459
QY 1141 atgttcaatcacgaaatctctgattctcaagctgcccacagagctgcgagcacaagct 1200
Db 827458 ATGTTCAATACGGAAATCTCTGATCTCAAGTGCCTCAACAGGAGCTCGCAGCACAACT 827399
QY 1201 agagcagcgaagccgctggagatgacagtgctgctgcagcgtgcgagatgctcagaaa 1260
Db 827398 AGAGCAGCGAAAGCCGCTGGAGATGACAGTGTGCTGAGCGCTGGCAGATGCTCAGAAA 827339
QY 1261 gctttagaagcgtctagtaaaagctggcaaacagcagcagcagcagcagcagcagcag 1320
Db 827338 GCTTTAGAAGCGGCTCTAGGTAAGCTGGGCAACAGCGGATCTACTCAATGCTTTAGGA 827279
QY 1321 cagatcgctcttctgctgctg-ttgtgagcgaggag-tcctcccgctgcagcaagttctat 1378

QY 1 atgtttaatcctattgtccagctcctatagacgaacagaacgacacaccccgagat 60
Db 828598 ATGTTTAAATCTATTTGTCAGGTCCTATAGACGAACAGACACACCTCCCGCAGAT 828539
QY 61 ctcttctcgaagattggagcgtgagtgacgaaataaagagtgcggaagctcaagaata 120
Db 828538 CTTTCTGCTCAAGGATTGGAGCGAGTGCAGCAATTAAGAGTGGGAAGCTCAAGAATA 828479
QY 121 gcaggtcggaagctaaagatctaaagaccgattctgtagagcagatgagcattc 180
Db 828478 GCAGGTGGGAAGCTAAAGCCTTAAAGATCTAAGACCACTGTTGAGAGCGATGGAGATC 828419
QY 181 ttgcgttctcagtgaaatgctctcctagctgctgagcagataagctgggtattgctctagt 240
Db 828418 TTGCGTCTGAGTAATGCTCTCTCATGAGTCTGCAGATAAGCTGGGTATGCTTCTAGT 828359

Db 827278 CAGATCGCTTCGCTGCTGTTTGTGAGCCAGGAGTTCCTCCCGTCGAGCAAGTTCAT 827219

QY 1379 qq 1380

Db 827218 AG 827217

RESULT 4

AAD02065
ID AAD02065 standard: DNA: 1456 bp.

XX
AC: AAD02065:

26-MAR-2001 (first entry)

3'-truncated *Chlamydia pneumoniae* 76 kDa protein encoding DNA.

76 kDa protein; bactericidal; diagnosis; prevention;
 pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 acute respiratory disease; cough; sore throat; hoarseness; fever;
 vaccine; immunisation; treatment; truncation mutant; ds.

OS Chlamydia pneumoniae.

OS Synthetic.

Location/Qualifiers	Key
101..1456	CDS

```
FT
000
/*tag= a
```

$$\text{FT} = \frac{\text{FT}}{\text{product} =}$$

FT
76KDa protein"

FT /note= "The co

$$\frac{\partial F_T}{\partial y}$$

XX PN WO200066739-A2

XX
PD
09-NOV-2000XX
XX
-00007 AON 60

PF 03-MAY-2000; 2000WO-CA00511.

XX

PR 03-MAY-1999; 99US-0132270.

PR 30-JUN-1999; 99US-0141276.
yy

XX
XX
PA / AVET \ AVENTIS PACTEUR LTDPA (AVET) AVENTIS PAST
XX
XX

PI Murdin AD, Oomen RP

XX
1954
1955
1956
1957
1958
1959
1960
1961
1962
1963
1964
1965
1966
1967
1968
1969
1970
1971
1972
1973
1974
1975
1976
1977
1978
1979
1980
1981
1982
1983
1984
1985
1986
1987
1988
1989
1990
1991
1992
1993
1994
1995
1996
1997
1998
1999
2000
2001
2002
2003
2004
2005
2006
2007
2008
2009
2010
2011
2012
2013
2014
2015
2016
2017
2018
2019
2020
2021
2022
2023
2024
2025
2026
2027
2028
2029
2030
2031
2032
2033
2034
2035
2036
2037
2038
2039
2040
2041
2042
2043
2044
2045
2046
2047
2048
2049
2050
2051
2052
2053
2054
2055
2056
2057
2058
2059
2060
2061
2062
2063
2064
2065
2066
2067
2068
2069
2070
2071
2072
2073
2074
2075
2076
2077
2078
2079
2080
2081
2082
2083
2084
2085
2086
2087
2088
2089
2090
2091
2092
2093
2094
2095
2096
2097
2098
2099
2100
2101
2102
2103
2104
2105
2106
2107
2108
2109
2110
2111
2112
2113
2114
2115
2116
2117
2118
2119
2120
2121
2122
2123
2124
2125
2126
2127
2128
2129
2130
2131
2132
2133
2134
2135
2136
2137
2138
2139
2140
2141
2142
2143
2144
2145
2146
2147
2148
2149
2150
2151
2152
2153
2154
2155
2156
2157
2158
2159
2160
2161
2162
2163
2164
2165
2166
2167
2168
2169
2170
2171
2172
2173
2174
2175
2176
2177
2178
2179
2180
2181
2182
2183
2184
2185
2186
2187
2188
2189
2190
2191
2192
2193
2194
2195
2196
2197
2198
2199
2200
2201
2202
2203
2204
2205
2206
2207
2208
2209
2210
2211
2212
2213
2214
2215
2216
2217
2218
2219
2220
2221
2222
2223
2224
2225
2226
2227
2228
2229
2230
2231
2232
2233
2234
2235
2236
2237
2238
2239
2240
2241
2242
2243
2244
2245
2246
2247
2248
2249
2250
2251
2252
2253
2254
2255
2256
2257
2258
2259
2260
2261
2262
2263
2264
2265
2266
2267
2268
2269
2270
2271
2272
2273
2274
2275
2276
2277
2278
2279
2280
2281
2282
2283
2284
2285
2286
2287
2288
2289
2290
2291
2292
2293
2294
2295
2296
2297
2298
2299
2300
2301
2302
2303
2304
2305
2306
2307
2308
2309
2310
2311
2312
2313
2314
2315
2316
2317
2318
2319
2320
2321
2322
2323
2324
2325
2326
2327
2328
2329
2330
2331
2332
2333
2334
2335
2336
2337
2338
2339
2340
2341
2342
2343
2344
2345
2346
2347
2348
2349
2350
2351
2352
2353
2354
2355
2356
2357
2358
2359
2360
2361
2362
2363
2364
2365
2366
2367
2368
2369
2370
2371
2372
2373
2374
2375
2376
2377
2378
2379
2380
2381
2382
2383
2384
2385
2386
2387
2388
2389
2390
2391
2392
2393
2394
2395
2396
2397
2398
2399
2400
2401
2402
2403
2404
2405
2406
2407
2408
2409
2410
2411
2412
2413
2414
2415
2416
2417
2418
2419
2420
2421
2422
2423
2424
2425
2426
2427
2428
2429
2430
2431
2432
2433
2434
2435
2436
2437
2438
2439
2440
2441
2442
2443
2444
2445
2446
2447
2448
2449
2450
2451
2452
2453
2454
2455
2456
2457
2458
2459
2460
2461
2462
2463
2464
2465
2466
2467
2468
2469
2470
2471
2472
2473
2474
2475
2476
2477
2478
2479
2480
2481
2482
2483
2484
2485
2486
2487
2488
2489
2490
2491
2492
2493
2494
2495
2496
2497
2498
2499
2500
2501
2502
2503
2504
2505
2506
2507
2508
2509
2510
2511
2512
2513
2514
2515
2516
2517
2518
2519
2520
2521
2522
2523
2524
2525
2526
2527
2528
2529
2530
2531
2532
2533
2534
2535
2536
2537
2538

DR WPI; 2000-687542/67.

DR P-PSDB; AAY71956.

[illegible]

PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae

XX
user for vaccinating

PS Claim 2c; Page 102-104; 112pp; English.

XX
XX

CC The present sequence is a DNA coding for 3'-truncated

CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa

CC protein is used in the
CC diagnosis prevention and treatment of C

CC diagnosis, prevention and treatment of *C. pneumoniae*
CC infections (e.g. pneumonia, upper respiratory

cc infections (e.g. pneumonia, upper respiratory tract disease, bronchitis, sinusitis and acute respiratory

Query Match

Query Match 91.68; Score 1347; DB 21; Length 1456;

Best Local Similarity 99.6%; Pred. No. 0;

QY	1	atggttaattccttattggttcacaggtcctatagacgaacacagaacgcacacctccgcagat	60
DB			
QY	61	ctttctgctcaaggattggaggcgaatgcagcaaatgaagtgcggaagctcaagaata	120
DB			
QY	161	ctctctgcacaggatttggaggcagtcgacaaataagagtgcggaagctcaagaata	220
DB			
QY	121	gcaggctgcgaagctaaagctaaagaactaaagaccgattctgtagagcgatggagcattc	180
DB			
QY	221	gcaggctgcgaagctaaagcctaaagaactaaagaccgattctgtagagcgatggagcattc	280
DB			
QY	181	ttgcttctgcagtgaaatgctctctatgagctggcagataaagctgggtattgcttctagt	240
DB			
QY	281	ttgcttctgcagtgaaatgctctctatgagctggcagataaagctgggtattgcttctagt	340
DB			
QY	241	aacagctcgtcttctactagcagatctgcagacgtggactcaacgacgagcgaccgacct	300
DB			
QY	341	aacagctcgtcttctactagcagatctgcagacgtggactcaacgacgagcgaccct	400
DB			
QY	301	acgctctctccaccacgctctgatgattataagactcaagcgcgaacagctctacgatact	360
DB			
QY	401	acgctctctccaccacgcttggatgattataagactcaagcgcgaacagctctacgatact	460
DB			
QY	361	atctttactctcaacatcactagctgacatacagctgcttttggctgagctccacgattgct	420
DB			
QY	461	atctttactctcaacatcactagctgacatacagctgcttttggctgagctccacgattgct	520
DB			
QY	421	gtcactaaataaaggatacagcgctactgatgaggaacccgcaactcgtctggagtggtg	480
DB			
QY	521	gtcactaaataaaggatacagcgctactgatgaggaacccgcaactcgtctggagtggtg	580
DB			
QY	481	gaactaaagaattgcgagtgcgaattaaagtgtggcgcgaataatcacagaatttagcgaaatat	540
DB			
QY	581	gaactaaagaattgcgagtgcgaattaaagtgtggcgcgaataatcacagaatttagcgaaatat	640
DB			
QY	541	gcttcggataacacgaagcattcttgactcttttagttaaactgactctctctcgacctctta	600
DB			
QY	641	gcttcggataacacgaagcattcttgactcttttagttaaactgactctctctcgacctctta	700
DB			
QY	601	cagactgctcttctccaattctgtagcaaacataacaaagcagctgagcttcttaagaag	660
DB			
QY	701	cagctgctcttctccaattctgtagcaaacataacaaagcagctgagcttcttaagaag	760
DB			
QY	661	atgcaagataacccagtagtccacagggaacgcgtcgaattgctcaactcttagttgct	720
DB			
QY	761	atgcaagataacccagtagtccacagggaacgcgtcgaattgctcaactcttagttgct	820
DB			
QY	721	cagacagatgctacgcgcacacagatagagaagaatggaatgcgattggggatgcataat	780
DB			
QY	821	caqacagatgctacagcgacacagatagagaagaatggaatgcgattggggatgcataat	880
DB			
QY	781	tttgacgagcagaacgcctagtggaactgtagaaaaatgctaaatctataacacgataaagc	840
DB			
QY	881	tttgacgagcagaacgcctagtggaactgtagaaaaatgctaaatctataacacgataaagc	940
DB			
QY	841	aacatagattcagctaaagcagcaactcgtcgtctaagacacacaaaagctgaagctcag	900
DB			
QY	941	aacatagattcagctaaagcagcaactcgtcgtctaagacacacaaaagctgaagctcag	1000
DB			
QY	901	aaaaagtctcccgactctccaatcttctcaagaagcggaacaaaatggttaatacaggtctgag	960
DB			
QY	1001	aaaaagtctcccgactctccaatcttctcaagaagcggaacaaaatggttaatacaggtctgag	1060
DB			
QY	961	aaagattcttaaaaaatatcaaacctgcagatggttctgatgttccaaatccaggaaactaca	1020
DB			
QY	1061	aaaga tcttaaaaaatatcaaacctgcagatggttctgatgttccaaatccaggaaactaca	1120
DB			
QY	1021	gttgagggtctcaagcaacaaaggaatgattggttagtattcgtgttctcatgctgtta	1080
DB			
QY	1121	gttgagggtctcaagcaacaaaggaatgattggttagtattcgtgttctcatgctgtta	1180
DB			
QY	1081	gatgatgctgaaaaatgagaccgcttccattttgatgtctggtgttctgcagatgattccac	1140

Db 1181 gatgatcgtgaaatgagacgcttcacatttgatgctggtttcgtcagatgattcac 1240
 Qy 1141 atgttcaatcagcaaatcctgattctcgaagctgcccacagagagctcgagcaacaagct 1200
 Db 1241 atgttcaatcagcaaatcctgattctcgaagctgcccacagagagctcgagcaacaagct 1300
 Qy 1201 agagcagcgaagccgctgagatgacagtgctcagcgcctgagcagctcgagcaaaa 1260
 Db 1301 agagcagcgaagccgctgagatgacagtgctcagcgcctgagcagctcgagcaaaa 1360
 Qy 1261 gcttttagaagcgcctcgaagtggaagctggcaacaacagggcactactcaatgctttggga 1320
 Db 1361 gcttttagaagcgcctcgaagtggaagctggcaacaacagggcactactcaatgcttttagga 1420
 Qy 1321 cagatcgcttctgctgctgtgtgagcgcaggagt 1355
 Db 1421 cagatcgcttctgctgctgtgtgagcgcaggagt 1455
 RESULT 5
 AAD02064
 ID AAD02064 standard; DNA; 1852 BP.
 AC AAD02064;
 XX AAD02064;
 DT 26-MAR-2001 (first entry)
 XX
 DE 5'-truncated Chlamydia pneumoniae 76 kDa protein encoding DNA.
 DE
 XX
 KW 76 kDa protein; bactericidal; diagnosis; prevention;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;
 KW vaccine; immunisation; treatment; truncation mutant; ds.
 XX
 OS Chlamydia pneumoniae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1752
 FT /tag= a
 FT /product= "5'-truncated Chlamydia pneumoniae
 FT 76kDa protein"
 FT /transl_except= (pos:1489..1491, aa:lle)
 XX
 PN WC0200066739-A2.
 XX
 XX 09-NOV-2000.
 XX
 XX 03-MAY-2000; 2000WO-CA00511.
 XX
 PR 03-MAY-1999; 99US-0132270.
 PR 30-JUN-1999; 99US-0141276.
 XX
 XX (AVET) AVENTIS PASTEUR LTD.
 PA
 XX Murdin AD, Oomen RP, Wang J, Dunn P;
 PI
 XX WPI; 2000-687542/67.
 DR P-PSDB; AAY1955.
 XX
 XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 PT useful for vaccinating against Chlamydia infections -
 PT
 XX Claim 2b; Page 97-99; 11pp; English.
 PS
 XX The present sequence is a DNA coding for 5'-truncated
 CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
 CC protein is used in the
 CC diagnosis, prevention and treatment of C. pneumoniae
 CC infections (e.g. pneumonia, upper respiratory
 CC tract disease, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and

CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX Sequence 1852 BP; 578 A; 409 C; 407 G; 458 T; 0 other;
 SQ
 Query Match 78.6%; Score 1155.4; DB 21; Length 1852;
 Best Local Similarity 99.4%; Pred. No. 0;
 Matches 1170; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
 Qy 205 atgagtctggcagataagctgggtattgttcttagtaaacagctgcttcttactagcaga 264
 Db 1 atgagtctggcagataagctgggtattgttcttagtaaacagctgcttcttactagcaga 60
 Qy 265 tctgcagacgtgactcaacgacgacgacgacgacgacgacgacgacgacgacgacgacg 324
 Db 61 tctgcagacgtgactcaacgacgacgacgacgacgacgacgacgacgacgacgacgacg 120
 Qy 325 gattataagactcaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 384
 Db 121 gattataagactcaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 180
 Qy 385 gacatacaggtctgttggtagcctccagagctgctgacgacgacgacgacgacgacgacg 444
 Db 181 gacatacaggtctgttggtagcctccagagctgctgacgacgacgacgacgacgacgacg 240
 Qy 445 gctactgatgagaaacccgcaatcgtcgaggtgggaaactaagaatgcccagtcgaatt 504
 Db 241 gctactgatgagaaacccgcaatcgtcgaggtgggaaactaagaatgcccagtcgaatt 300
 Qy 505 aaagtggcgcgcaaatcagaaattacagaaattagcgaatattgcttcgataaaccagcattct 564
 Db 301 aaagtggcgcgcaaatcagaaattacagaaattagcgaatattgcttcgataaaccagcattct 360
 Qy 565 gactcttttaggtaaactgactcttcgacgcttccagccttcagagctgcttctccaatcgtga 624
 Db 361 gactcttttaggtaaactgactcttcgacgcttccagccttcagagctgcttctccaatcgtga 420
 Qy 625 gcaacaatacaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 684
 Db 421 gcaacaatacaacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacgacg 480
 Qy 685 gggaaaacgctgcaattgctcaatctttagtgatcagacagatgctacagcagacag 744
 Db 481 gggaaaacgctgcaattgctcaatctttagtgatcagacagatgctacagcagacag 540
 Qy 745 atagagaaagatggaatgcatgggagtgatgcatatttgcagagacagacgctagtgga 804
 Db 541 atagagaaagatggaatgcatgggagtgatgcatatttgcagagacagacgctagtgga 600
 Qy 805 gctgtagaaaaatgcttaaatcaacagtaataagcaacatagattcagctaaagcagca 864
 Db 601 gctgtagaaaaatgcttaaatcaacagtaataagcaacatagattcagctaaagcagca 660
 Qy 865 atcgtactgtaagacacacaaatagctgaagctcagaaaaagttcccgactctccaaatt 924
 Db 661 atcgtactgtaagacacacaaatagctgaagctcagaaaaagttcccgactctccaaatt 720
 Qy 925 cttcaagagcggcaacaaatgtaatacaggtcagaaagatctttaaataatcaaacct 984
 Db 721 cttcaagagcggcaacaaatgtaatacaggtcagaaagatctttaaataatcaaacct 780
 Qy 985 gcagatggttctgtatgttccaaatccaggaactacagttgaggtcctcaacacacagga 1044
 Db 781 gcagatggttctgtatgttccaaatccaggaactacagttgaggtcctcaacacacagga 840
 Qy 1045 actagtattgtagtattcgtgttccatgctgttagatgactgtaaaatgagaccgt 1104
 Db 841 agtagtattgtagtattcgtgttccatgctgttagatgactgtaaaatgagaccgt 900
 Qy 1105 tccattttagtctgggtttcgtcagatgattcactgttcaatcacggaataatcgtat 1164

Db 901 tccatttgatgctgggttctgcagatgattcacatgtttcaatacaggaatacctctgat 960
 Qy 1165 tctcaagtgcccaacagagactgcagacacaaagctagagcagcagaaagccgctggagat 1224
 Db 961 tctcaagtgcccaacagagactgcagacacaaagctagagcagcagaaagccgctggagat 1020
 Qy 1225 gacagtgctgctgcagcgtgcagatgctcagaaaagcttttagaagcgcctctaggtaaa 1284
 Db 1021 gacagtgctgctgcagcgtgcagatgctcagaaaagcttttagaagcgcctctaggtaaa 1080
 Qy 1285 gctgggcaacacagggcactactcaatgctttggacagatcgcttctgctgtgtgtg 1344
 Db 1081 gctgggcaacacagggcactactcaatgcttttagaagcgcctctaggtaaa 1140
 Qy 1345 agcgcagag-tctctccgctgcagcaagttctatgg 1380
 Db 1141 agcgcagagttctctccgctgcagcaagttctatag 1177

RESULT 6
 AAV16207
 ID AAV16207 standard; DNA; 150 BP.
 XX
 AC AAV16207;
 XX
 DT 28-MAY-1998 (first entry)
 XX
 DE Part of the gene encoding the 76 kDa cysteine rich OMP.
 XX
 KW Cysteine rich outer-membrane protein; OMP; 60 kDa OMP; 76 kDa OMP;
 KW hybridisation; amplification; assay; detection; ds.
 XX
 OS Chlamydia pneumoniae.
 XX
 PN WO9746709-A2.
 XX
 PD 11-DEC-1997.
 XX
 PF 04-JUN-1997; 97WO-US09673.
 XX
 PR 06-JUN-1996; 96US-0659473.
 XX
 PA (ABBO) ABBOTT LAB.
 XX
 PI Cerney MB;
 XX
 DR WPI; 1998-042215/04.
 XX
 PT Chlamydia pneumoniae derived oligonucleotides - used as primers or
 PT probes for specific and sensitive detection
 XX
 PS Disclosure; Page 24; 28pp; English.
 XX

CC The present sequence represents part of the gene encoding the 76 kDa
 CC cysteine rich outer-membrane protein (OMP) of Chlamydia pneumoniae.
 CC Novel oligonucleotides AAV16195-206 can function either as PCR primers
 CC or probes. They are used to detect C. pneumoniae in a sample.
 CC Oligonucleotides V161202-04 are specific for the present sequence.
 CC The oligonucleotides can be used in a hybridisation or amplification
 CC based assay for the detection of C. pneumoniae in a test sample.
 XX
 SQ Sequence 150 BP; 43 A; 34 C; 40 G; 33 T; 0 other;

Query Match 10.08; Score 146.8; DB 19; Length 150;
 Best Local Similarity 98.78; Pred. No. 1.1e-32;
 Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 Qy 366 tacctcaacatactagctgacacagctgctgttggtagcctccagagatgctgtcac 425
 Db 1 tacctcaacatactagctgacacagctgctgttggtagcctccagagatgctgtcac 60
 Qy 426 taataaaggatacagcggctactgactgatgaggaacacgcaatcgctgcggagtgggaac 485

Db 61 taataaaggatacagcggctactgactgatgaggaacacgcaatcgctgcggagtgggaac 120
 Qy 486 taagaatgccgatgcaattaaagtggcgc 515
 Db 121 taagaatgccgatgcaattaaagtggcgc 150

RESULT 7
 AAF58252/C
 ID AAF58252 standard; DNA; 936 BP.
 XX
 AC AAF58252;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D1835.
 XX
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX
 OS Synthetic.
 XX
 PN WO200107665-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI Umek RM;
 XX
 DR WPI; 2001-159728/16.
 XX
 PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.68; Score 83; DB 22; Length 936;
 Best Local Similarity 1.2%; Pred. No. 7.9e-14;
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;
 Qy 482 aaactaagaatgcgagatgcaattaaagtggcgcgcaataacaggaatagcgaatag 541
 Db 750 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 691
 Qy 542 ctgcgatacaacagcagatcttgaactctttagttaaactgaactcttcctcactcttac 601
 Db 690 GWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 631
 Qy 602 agactgctctctccaatctgtagcaacaatacaacagcagctgagcttctttaaagaga 661
 Db 630 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 571
 Qy 662 tgcagaataacccagtagtccacaggaacgcctgcaatgctgcaatctttagttagtc 721
 Db 570 WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW WWWWWW 511

OS Synthetic.
 PN WO200107665-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI Umek RM;
 XX
 DR WPI; 2001-159728/16.
 XX
 PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 127; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 SQ Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.6%; Score 83; DB 22; Length 936;
 Best Local Similarity 1.2%; Pred. No. 7.9e-14;
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatgccgatgcaattaaagtgtggtgcgcgaataacagaaattagcgaataatg 541
 DB 750 WWWWWW
 QY 542 ctctcgataaccagcagattcttgactcttttagttaaactgactctcttcgacctttac 601
 DB 690 GWWWWW
 QY 602 agactgctcttctcccaatctgtagcaacaatacaaacagcagctgagctctttaaagaga 661
 DB 630 WWWWWW
 QY 662 tgcagataaccagtagtccagggaaacgcctgcattgctcaattctttagtgatc 721
 DB 570 WWWWWW
 QY 722 agacagatgctacagcacagatagagaagatggaatgcgattgggagatcatatt 781
 DB 510 WWWWWW
 QY 782 ttgcaggacagacgcagtagtgagctagaaatgtcaaatcttaataacagataagca 841
 DB 450 WWWWWW
 QY 842 acatagattcagctaaagcagcaatcgctactgtcttaagacacaaatagctgaagctcaga 901
 DB 390 WWWWWW
 QY 902 aaaagtcccgactctcaattcttcaagaagcgaacaaatgtaatacagcgtgaga 961
 DB 330 WWWWWW
 QY 962 agagattctaaataatcaaacctgcagatggtcttgatgttcccaatccaggaactacag 1021
 DB 270 WWWWWW

QY 1022 ttggaggtcccaagcaacaagaagtagtattggttagtattgtttccatgctgttag 1081
 DB 210 WWWWWW
 QY 1082 atgatgtgaaaaatgagaccgcttccatttgcattgtctgggttcgtcagatgattcaca 1141
 DB 150 WWWWWW
 QY 1142 tgttcaatacgcgaaaaatcctgattctcaa 1170
 DB 90 WWWWWW

RESULT 10
 AAF58259/c
 ID AAF58259 standard; DNA; 936 BP.
 XX
 AC AAF58259;
 XX
 DT 24-APR-2001 (first entry)
 XX
 DE Oligonucleotide D2004.
 XX
 KW Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 OS Synthetic.
 PN WO200107665-A2.
 XX
 PD 01-FEB-2001.
 XX
 PF 26-JUL-2000; 2000WO-US20476.
 XX
 PR 26-JUL-1999; 99US-0145695.
 PR 17-MAR-2000; 2000US-0190259.
 XX
 PA (CLIN-) CLINICAL MICRO SENSORS INC.
 XX
 PI Umek RM;
 XX
 DR WPI; 2001-159728/16.
 XX
 PT Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX
 PS Example 6; Page 128; 159pp; English.
 XX
 CC The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX
 SQ Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 5.6%; Score 83; DB 22; Length 936;
 Best Local Similarity 1.2%; Pred. No. 7.9e-14;
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatgccgatgcaattaaagtgtggtgcgcgaataacagaaattagcgaataatg 541
 DB 750 WWWWWW
 QY 542 ctctcgataaccagcagattcttgactcttttagttaaactgactctcttcgacctttac 601
 DB 690 GWWWWW
 QY 602 agactgctcttctcccaatctgtagcaacaatacaaacagcagctgagctctttaaagaga 661
 DB 630 WWWWWW

DE Oligonucleotide D1876.
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS
XX WO200107665-A2.
PN
XX
XX 01-FEB-2001.
PD
XX 26-JUL-2000; 2000WO-US20476.
PF
XX 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX Umek RM;
PI WPI; 2001-159728/16.
DR
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
PT
XX
XX Example 6; Page 127; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;
SQ

Query Match 5.6%; Score 83; DB 22; Length 938;
Best Local Similarity 1.2%; Pred. No. 7.9e-14;
Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatccgatgcaatgaattggtggcgcaaatattacagaattagcgaaatag 541
DB 750 WWWWWW... 691
QY 542 ctccggataaccaagcattctgactcttaggttaaaactgactcttcgaccttac 601
DB 690 GHHWWW... 631
QY 602 agactgctcttctccaatctgtagcaaaataacaaagcagctgagcttcttaagaga 661
DB 630 WWWWWW... 571
QY 662 tgaagataaaccagtagtccagggagaaacgcctgcaattgctcaatttagttgatc 721
DB 570 WWWWWW... 511
QY 722 agacagtgctcacgcacagatagagaaagatggaatgcgattgggagatcatatt 781
DB 510 WWWWWW... 451
QY 782 ttgcaggacagacgctagtgagctgtagaaatctgtaaatcttaataacagtataagca 841
DB 450 WWWWWW... 391
QY 842 acatagattcagtaaacgacgaatcgctactctgtaacacaaatagctgaagctcaga 901
DB 390 WWWWWW... 331
QY 902 aaaagtctcccgactctccaattcttcaagaagcggaaacaaatgtaatacaggtcaga 961
DB 330 WWWWWW... 271

QY 962 aagatcttaaaaaataatcaaacctgcagatggtctctgattgttccaaatccaggaaactacag 1021
DB 270 WWWWWW... 211
QY 1022 ttggagggtcccaagcaacaaaggaagtagtattgtagttattcgtgttccatgctgttag 1081
DB 210 WWWWWW... 151
QY 1082 atgatgcgaaatgagaccgcttcatttgcatttgcgtggttgcgtcagatgattcaca 1141
DB 150 WWWWWW... 91
QY 1142 tttcaatacaggaataatcctgattctcaa 1170
DB 90 WWWWWW... 62

RESULT 13
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX
AC AAF58252;
XX
DT 24-APR-2001 (first entry)
XX
DE Oligonucleotide D1835.
XX
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX
OS Synthetic.
XX
PN WO200107665-A2.
XX
PD 01-FEB-2001.
XX
PF 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;
XX
DR WPI; 2001-159728/16.
XX
PT Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
PS Example 6; Page 127; 159pp; English.
XX
CC The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.5%; Score 80.6; DB 22; Length 936;
Best Local Similarity 1.2%; Pred. No. 3.9e-13;
Matches 8; Conservative 402; Mismatches 281; Indels 0; Gaps 0;

QY 483 aactaagaatgcgcatgcaattaaagtggcgcgcaaatattacagaattagcgaaatagc 542
DB 1 WWWWWW... 50
QY 543 ttccgataaccaagcgaattctttaggttaactctttagtactcttccttcgacctttaca 602

[illegible]

RESULT 14
AAAF58254
ID AAAF58254 standard; DNA; 936 BP.

XX AAF58254;
AC

XX

DT 24-APR-2001 (first entry)

XX
DE Oligonucleotide D1875.

XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
KW

XX Synthetic.

XX
PN WO200107665-A2.

XX 01-FEB-2001.
PD

XX
PF 26-JUL-2000; 2000WO-US20476.

XX
PR 26-JUL-1999; 99US-0145695.

PR 17-MAR-2000; 2000US-0190259.
XX

PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX
PI Umek RM;

XX
DR WPI; 200

XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface -
XX
XX Example 6; Page 127; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETW) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
SQ

Query Match	5.5%	Score 80.6;	DB 22;	Length 936;
Best Local Similarity	1.2%	Pred. NO. 3.9e-13;		
8. Conservative Matches	402;	Mismatches 281;		
Gaps	0;	Indels	0;	

Qy	483	aactaagaatgccgatgcaattaaagttg	cgcgcaaaattacagaattagcgaatatgc	542
Db	1			60
Qy	543	ttcggataaccaagcgattctt	gactcttaggtaaactgacttcttcgacctttaca	602
Db	61			120
Qy	603	gactgctcttccaatctgtagcaaacaa	tatacaaacgagctgagctctttaagagat	662
Db	121			180
Qy	663	gcaagataaccagctagtc	ccagggaacgcctgcaattgctcaatttttagttgata	722
Db	181			240
Qy	723	gacagatgctacagcacacagatagaa	agatggaatgcgattggggatgcataatt	782
Db	241			300
Qy	783	tgcaggacagacgctagtgagctgt	agaaaatgtaactaataacagataagca	842
Db	301			360
Qy	843	catagattcagctaaagcagcaatcgct	tactctgaagcacacaaatagctgaagctcagaa	902
Db	361			420
Qy	903	aaagtcccgactctccaatcttcaaga	agcgaaacaaatggtaatacagggctgagaa	962
Db	421			480
Qy	963	agatctaaaaatatcaaacctgcagatg	ttctgatgttccaaatccaggcaactacagt	1022
Db	481			540
Qy	1023	tggaggctccaagcaacaaaggtagta	tatggtagtattcgtgtttccatgctgttaga	1082
Db	541			600
Qy	1083	tgatgctaaaaatgagacgcgttccatt	tttgatgctctgctgagatgattcacat	1142
Db	601			660
Qy	1143	gttcaatacggaaaatcctgattctca	agct	1173
Db	661			691

RESULT 15
AAF58257
ID AAF58

AAF58257
ID AAF58257 standard; DNA; 936 BP.

XX AAF58257;
AC 24-APR-2001 (first entry)
DT XX
DE Oligonucleotide D1954.
XX Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
XX Synthetic.
OS
XX WO2000107665-A2.
PN
XX
PD
XX
PF 01-FEB-2001.
XX 26-JUL-2000; 2000WO-US20476.
XX
PR 26-JUL-1999; 99US-0145695.
PR 17-MAR-2000; 2000US-0190259.
XX
XX (CLIN-) CLINICAL MICRO SENSORS INC.
PA
XX
PI Umek RM;
XX
XX WPI; 2001-159728/16.
DR
XX
XX Nucleic acids containing electron-transfer group, useful as labels in
PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
PT a single surface
XX
XX Example 6; Page 127; 159pp; English.
PS
XX The present invention relates to a composition comprising two nucleic
CC acids each containing an electron-transfer group (ETM) having
CC different redox potentials. The invention is used for electronic
CC detection of nucleic acids, especially of substitutions (mismatches)
CC and single-nucleotide polymorphisms, e.g. for genotyping,
CC monitoring gene expression.
XX
XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 5.5%; Score 80.6; DB 22; Length 936;
Best Local Similarity 1.2%; Pred. No. 3.9e-13;
Matches 8; Conservative 402; Mismatches 281; Indels 0; Gaps 0;

QY 483 aactaagaatgccgatgcaattaaagtggcgcgcaaatcagaaattagcgaaatagc 542
DB 1
QY 543 ttccgataaccaagcgattcttgactcttttaggttaaacgactcttcctcgacctttaca 602
DB 61
QY 603 gactgctcttccaatctgtagcaacaataacaaagcagctgactctttaaagagat 662
DB 121
QY 663 gcaagataaccagtagtcccgagggaacgcctgcaattgctcaatttttagttgatca 722
DB 181
QY 723 gacagatgctacagcgacacagatagagaagaatggaatggattggggatgcatttt 782
DB 241
QY 783 tgcaggacagaacgcgtagtggagctgtagaaaaatgctaaataacacagataagcaa 842
DB 301
QY 843 catagattcagctaaagcagcaaatcgctactgctaagacacaaatagctgaagctcagaa 902

DB 361
QY 903
DB 421
QY 963
DB 481
QY 1023
DB 541
QY 1083
DB 601
QY 1143
DB 661

Search completed: February 7, 2002, 19:59:45
Job time: 15099 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:53:37 ; Search time 243.49 Seconds
(without alignments)
1367.295 Million cell updates/sec

Title: US-09-391-606-3
Perfect score: 1470
Sequence: 1 atggttaatctattgttc.....accatcatcatcatcat 1470

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues
Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146.8	10.0	150	4	US-08-659-473-8
2	65	4.4	816	4	US-09-411-329C-19
3	65	4.4	1373	4	US-09-411-329C-20
4	65	4.4	1620	4	US-09-411-329C-12
5	65	4.4	1620	4	US-09-411-329C-13
6	54.4	3.7	7218	1	US-08-232-463-14
7	50.6	3.4	125	2	US-08-350-260A-41
8	50.6	3.4	2085	1	US-08-072-070-1
9	50.6	3.4	2085	1	US-08-465-746-1
10	50.6	3.4	2085	1	US-08-214-164-1
11	50.6	3.4	2085	1	US-08-469-434-1
12	50.6	3.4	2085	1	US-08-214-222-1
13	50.6	3.4	2085	2	US-08-467-852A-1
14	50.6	3.4	2085	2	US-08-468-718-1
15	50.6	3.4	2085	2	US-08-246-636-1
16	50.6	3.4	2085	2	US-08-247-431A-1
17	50.6	3.4	2085	2	US-08-319-795-1
18	50.6	3.4	2085	2	US-08-468-985-1
19	50.6	3.4	2086	3	US-08-312-949-1
20	50.6	3.4	2086	3	US-08-446-201-2
21	50.4	3.4	217	2	US-08-958-201-17
22	49.8	3.4	913	2	US-08-661-052-13
23	49.8	3.4	913	2	US-09-188-082-13
24	49.8	3.4	1679	2	US-08-661-052-15
25	49.8	3.4	1679	2	US-09-188-082-15
26	49	3.3	907	3	US-09-184-658-9
27	48.8	3.3	216	2	US-08-652-816A-40

28	43.6	3.0	6744	1	US-08-119-125A-2	Sequence 2, Appli
29	39.8	2.7	72	2	US-08-273-146-36	Sequence 36, Appl
30	39.8	2.7	123	2	US-08-273-146-38	Sequence 38, Appl
31	39.8	2.7	132	2	US-08-273-146-40	Sequence 40, Appl
32	39.8	2.7	274	2	US-08-448-418-96	Sequence 96, Appl
33	39.8	2.7	883	3	US-09-184-658-7	Sequence 7, Appli
34	39.6	2.7	126	1	US-08-211-202-131	Sequence 131, App
35	39.6	2.7	162	1	US-08-211-202-132	Sequence 132, App
36	39.6	2.7	162	2	US-08-448-418-94	Sequence 94, Appl
37	39.6	2.7	640	3	US-08-961-083-33	Sequence 33, Appl
38	39.6	2.7	4403	5	PCT-US93-01901-30	Sequence 30, Appl
39	39.4	2.7	840	4	US-09-260-527-4	Sequence 4, Appli
40	39.4	2.7	3337	1	US-08-072-610-1	Sequence 1, Appli
41	39.4	2.7	3337	2	US-08-719-822B-1	Sequence 1, Appli
42	39.4	2.7	3337	4	US-09-092-458-1	Sequence 1, Appli
43	38	2.6	6896	2	US-08-627-151A-6	Sequence 6, Appli
44	37.4	2.5	1920	1	US-08-186-222-1	Sequence 1, Appli
45	37.4	2.5	5361	4	US-08-973-462-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-659-473-8
; Sequence 8, Application US/08659473
; Patent No. 6210876
; GENERAL INFORMATION:
; APPLICANT: M. B. Cerney
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.0.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,473
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul D. Yasger
; REGISTRATION NUMBER: 37,477
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708/937-2341
; TELEFAX: 708/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA (C. pneumoniae)
US-08-659-473-8

Query Match 10.0%; Score 146.8; DB 4; Length 150;
Best Local Similarity 98.7%; Pred. No. 3.2e-34;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 366 tacctcaacatcatcactagctgacatcacagctcttttgtagcctccagatgctgtcac 425
|||||
DB 1 TACCTCAACATCCTAGCTAGCTGACATACAGGCTGCTTTGGTGAGCCTCCAGATGCTGTAC 60

Qy 426 taatataaagatcacagcggtactatgatgaggaacccgcaatcgctcggagtggaac 485
 Db 61 TAATATAAGGATACAGCGGTACTGATGAGGAACCCGCAATCGCTCGGTGTGGGAAC 120
 Qy 486 taagaatccgatcaattaaagtggcg 515
 Db 121 TAAGATCCGATCGATGATTAAGTTGGCG 150

RESULT 2
 US-09-411-329C-19
 ; Sequence 19, Application US/09411329C
 ; Patent No. 6261820
 ; GENERAL INFORMATION:
 ; APPLICANT: Boone, Thomas
 ; APPLICANT: Li, Huimin
 ; APPLICANT: Mann, Michael
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
 ; FILE REFERENCE: A-596
 ; CURRENT APPLICATION NUMBER: US/09/411,329C
 ; CURRENT FILING DATE: 1999-10-01
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 19
 ; LENGTH: 816
 ; TYPE: DNA
 ; ORGANISM: Agkistrodon contortrix
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(816)
 ; OTHER INFORMATION: Fragment of fibrolase of Agkistrodon contortrix
 US-09-411-329C-19

Query Match 4.4%; Score 65; DB 4; Length 816;
 Best Local Similarity 100.0%; Pred. No. 1.4e-09;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaataactctcagaagaggtcgaatagcgcgcgtcgaccatcatcatc 1465
 Db 646 tagaacaataactctcagaagaggtcgaatagcgcgcgtcgaccatcatcatc 705
 Qy 1466 atcat 1470
 Db 706 atcat 710

RESULT 3
 US-09-411-329C-20
 ; Sequence 20, Application US/09411329C
 ; Patent No. 6261820
 ; GENERAL INFORMATION:
 ; APPLICANT: Boone, Thomas
 ; APPLICANT: Li, Huimin
 ; APPLICANT: Mann, Michael
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
 ; FILE REFERENCE: A-596
 ; CURRENT APPLICATION NUMBER: US/09/411,329C
 ; CURRENT FILING DATE: 1999-10-01
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 20
 ; LENGTH: 1373
 ; TYPE: DNA
 ; ORGANISM: Agkistrodon contortrix
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(1373)
 ; OTHER INFORMATION: Fragment of fibrolase of Agkistrodon contortrix
 US-09-411-329C-20

Query Match 4.4%; Score 65; DB 4; Length 1373;
 Best Local Similarity 100.0%; Pred. No. 1.9e-09;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaataactctcagaagaggtcgaatagcgcgcgtcgaccatcatcatc 1465
 Db 1203 tagaacaataactctcagaagaggtcgaatagcgcgcgtcgaccatcatcatc 1262
 Qy 1466 atcat 1470
 Db 1263 atcat 1267

RESULT 4
 US-09-411-329C-12
 ; Sequence 12, Application US/09411329C
 ; Patent No. 6261820
 ; GENERAL INFORMATION:
 ; APPLICANT: Boone, Thomas
 ; APPLICANT: Li, Huimin
 ; APPLICANT: Mann, Michael
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
 ; FILE REFERENCE: A-596
 ; CURRENT APPLICATION NUMBER: US/09/411,329C
 ; CURRENT FILING DATE: 1999-10-01
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 12
 ; LENGTH: 1620
 ; TYPE: DNA
 ; ORGANISM: Agkistrodon contortrix
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(1620)
 ; OTHER INFORMATION: Complementary (sense) strand of antisense strand (See SEQ ID
 ; OTHER INFORMATION: 3
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(1620)
 ; OTHER INFORMATION: Coding sequence of native pro-fibrolase of Agkistrodon contor
 US-09-411-329C-12

Query Match 4.4%; Score 65; DB 4; Length 1620;
 Best Local Similarity 100.0%; Pred. No. 2e-09;
 Matches 65; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1406 tagaacaataactctcagaagaggtcgaatagcgcgcgtcgaccatcatcatc 1465
 Db 1412 tagaacaataactctcagaagaggtcgaatagcgcgcgtcgaccatcatcatc 1471
 Qy 1466 atcat 1470
 Db 1472 atcat 1476

RESULT 5
 US-09-411-329C-13/c
 ; Sequence 13, Application US/09411329C
 ; Patent No. 6261820
 ; GENERAL INFORMATION:
 ; APPLICANT: Boone, Thomas
 ; APPLICANT: Li, Huimin
 ; APPLICANT: Mann, Michael
 ; TITLE OF INVENTION: FIBRINOLYTICALLY ACTIVE POLYPEPTIDE
 ; FILE REFERENCE: A-596
 ; CURRENT APPLICATION NUMBER: US/09/411,329C
 ; CURRENT FILING DATE: 1999-10-01
 ; NUMBER OF SEQ ID NOS: 29
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 13
 ; LENGTH: 1620
 ; TYPE: DNA
 ; ORGANISM: Agkistrodon contortrix
 ; FEATURE:
 ; NAME/KEY: misc_feature
 ; LOCATION: (1)..(1620)
 ; OTHER INFORMATION: Complementary (antisense) strand of sense strand (See SEQ ID


```

RESULT      8
US-08-072-070-1
; Sequence 1, Application US/08072070
; Patent No. 5476929
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: Yother, Janet L
; APPLICANT: McDaniel, Larry S
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,070
; FILING DATE: 19930603
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:

```

```

RESULT          9
US-08-465-746-1
; Sequence 1, Application US/08465746
; Patent No. 5679768
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: Yother, Janet L
; APPLICANT: McDaniel, Larry S
; TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE
; TITLE OF INVENTION: PROTEIN A
; NUMBER OF SEQUENCES: 5

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: Suite 1203, 2001 Jefferson Davis Highway
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE: US/08/465.746
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/08/048.896
FILING DATE: US/07/656.773
APPLICATION NUMBER: US/07/835.698
FILING DATE: 12-FEB-1992
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 521-0378
TELEX: LURPAT WASHINGTON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Streptococcus pneumoniae
STRAIN: Rx1
IMMEDIATE SOURCE:
CLONE: JY4313
FEATURE:
NAME/KEY: intron
LOCATION: 1..2085
FEATURE:
NAME/KEY: CDS
LOCATION: join(127..1984)

US-08-465-746-1
Query Match 3.4%; Score 50.6; DB 1; Length 2085;
Best Local Similarity 49.8%; Pred. No. 4.3e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;
Qy 745 atagagaagatggaatgcgattgggagtcataatttgcaggacagaaagcgtagtggga 804
Db 472 ACAGACAAGCGCGAAGAACGACGACGACGACGACGACGACGACGACGACGACGACG 531
Qy 805 gctgtagaaaatgctaaatcctaataacagataaagcagataagcagataagcagcagca 864
Db 532 GAAGAGGCAAAACTAAATTTAACTGTTTCGAGCAATGGTAGTTCTTGAGCCAGCAG 591
Qy 865 atcgtactgtctgaagcagcaaaatgctgaagctcagaaaaagttcccca ---ctctcca 921
Db 592 TTGGCTGAGACTAAGAAAAATCAGAGAAGCTAAACAAAAGCACCAGACTTACTTAA 651
Qy 922 attcttcagaagcggcaaatggtaatacaggtcgtagaagatcttataaaatcaaa 981
Db 652 AAACTAGAGAGCTTAAAGCAAAATTTAGAGAGGCTGAGAAAAAGCTACTGAAGCCAAA 711
Qy 982 cctgcagatggtctgtgatttcccaatccaggaactacagttggaggctccaagcaaa 1041

Db 712 CAAAAAGTGATGCTGAAGAAGTCGCTCTCAAGCTAAATCGCTGAATTGAAAAATCAA 771
Qy 1042 ggaagtagtattg 1054
Db 772 GTTCATAGACTAG 784
RESULT 10
US-08-214-164-1
Sequence 1, Application US/08214164
Patent No. 5728387
GENERAL INFORMATION:
APPLICANT: BRILES, DAVID E.
APPLICANT: YOTTER, JANET L.
TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: Suite 1203, 2001 Jefferson Davis Highway
CITY: Arlington
STATE: Virginia
COUNTRY: U.S.A.
ZIP: 22202-0286
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/214.164
FILING DATE: 17-MAR-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/656.773
FILING DATE: 15-FEB-1991
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Berkstresser, Jerry W.
REGISTRATION NUMBER: 22,651
REFERENCE/DOCKET NUMBER: 6102-137
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 415-0810
TELEFAX: (703) 521-0813
TELEX: LURPAT WASHINGTON
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2085 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: intron
LOCATION: 1..1983
FEATURE:
NAME/KEY: CDS
LOCATION: 127..1983
US-08-214-164-1

Query Match 3.4%; Score 50.6; DB 1; Length 2085;
Best Local Similarity 49.8%; Pred. No. 4.3e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;
Qy 745 atagagaagatggaatgcgattgggagtcataatttgcaggacagaaagcgtagtggga 804
Db 472 ACAGACAAGCGCGAAGAACGACGACGACGACGACGACGACGACGACGACGACGACG 531
Qy 805 gctgtagaaaatgctaaatcctaataacagataaagcagataagcagataagcagcagca 864
Db 532 GAAGAGGCAAAACTAAATTTAACTGTTTCGAGCAATGGTAGTTCTTGAGCCAGCAG 591
Qy 865 atcgtactgtctgaagcagcaaaatgctgaagctcagaaaaagttcccca ---ctctcca 921

Db	592	TTGGCTGAGACATAGAAAAAATCAGAAGAGCTAAACAAAAAGCAGCAACTTACTATAA	651
Qy	922	attcttcaagaagcggaaacaaatggtatcacaggtgagaagatcttataaatatcaaa	981
Db	652	AAACTAGAGAAGCTAAAGCAAAATTAGAAGAGGCTGAGAAAAAGCTACTGAAGCCAAA	711
Qy	982	ctgcagatggcttcgatgttccaaatccagaaactcagttggaggtcccaagcaaaa	1041
Db	712	CAAAAAGTGATGCTGAGAAGTCGCTCCTCAAGCTAAATCGCTGAATTGGAAAAATCAA	771
Qy	1042	ggaagtagtattg	1054
Db	772	GTTCCATAGACTAG	784

RESULT 11
 US-08-469-434-1
 ; Sequence 1, Application US/08469434
 ; Patent No. 5753463
 ; GENERAL INFORMATION:
 ; APPLICANT: Briles, David E
 ; APPLICANT: Yother, Janet L
 ; APPLICANT: McDaniel, Larry S
 ; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
 ; TITLE OF INVENTION: PROTEIN
 ; NUMBER OF SEQUENCES: 6
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Shoemaker and Mattare, Ltd
 ; STREET: Sulite 1203, 2001 Jefferson Davis Highway
 ; CITY: Arlington
 ; STATE: Virginia
 ; COUNTRY: U.S.A.
 ; ZIP: 22202-0286
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/469,434
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/072,065
 ; FILING DATE: 03 JUNE 1993
 ; APPLICATION NUMBER: US/07/835,698
 ; FILING DATE: 12-FEB-1992
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/07/656,773
 ; FILING DATE: 15-FEB-1991
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (703) 415-0810
 ; TELEFAX: (703) 521-0378
 ; TELEX: LUKPAT WASHINGTON
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2085 base pairs
 ; TYPE: nucleic acid
 ; STRANDEDNESS: double
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; HYPOTHETICAL: NO
 ; ANTI-SENSE: NO
 ; ORIGINAL SOURCE:
 ; ORGANISM: Streptococcus pneumoniae
 ; STRAIN: Rx1
 ; IMMEDIATE SOURCE:
 ; CLONE: JY2008
 ; FEATURE:
 ; NAME/KEY: intron
 ; LOCATION: 1..2085
 ; FEATURE:

ADDRESSEE: Shoemaker and Mattare, Ltd
STREET: Suite 1203, 2001 Jefferson Davis Highway

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/468,718

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/08/072,068

FILING DATE: 03 JUNE 1993

APPLICATION NUMBER: US/07/835,698

FILING DATE: 12-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/07/656,773

FILING DATE: 15-FEB-1991

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 521-0378

TELEX: LUKPAT WASHINGTON

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2085 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: Streptococcus pneumoniae

STRAIN: Rxl

IMMEDIATE SOURCE:

CLONE: JY2008

FEATURE:

NAME/KEY: Intron

LOCATION: 1..2085

FEATURE:

NAME/KEY: CDS

LOCATION: Join(127..1983, 1987..1992, 1996..2007, 2011

LOCATION: ..2025, 2029..2031, 2035..2085)

US-08-468-718-1

Query Match 3.4%; Score 50.6; DB 2; Length 2085;
Best Local Similarity 49.8%; Pred. No. 4.3e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

QY 745 atagagaagatgaaatgcgattgggtggtatatttgcagcagacagcgtagtga 804
Db 472 ACAGACAAACCCCAAAAGCAGCAGATAGATAGATAGAACTAAGAACCGAA 531
QY 805 gctgtagaagaatgctaaatcctaataacagctataagcaacatagattcagctaaagcaga 864
Db 532 GAAGAGCAAAACTAAATTTAATCTGTTGAGCAATGCTAGTTCCTGAGCCAGAGCAG 591
QY 865 atcgctactgtaagacacaaatagctgaagctcagaaaaagttcccca--ctctcca 921
Db 592 TTGGCTGAGACTAAGAAAAAATCAGAAAGCTTAACAAAAACACCACTACTATAA 651
QY 922 attctcaagaagcgggaacaaatggttaatacagctgagaaaaatcttaaaaaatacaaa 981
Db 652 AAACAGAGAGCTAAAGCAAAATTAGAAGAGGCTGAGAAAAAAGCTACTGAAGCCAAA 711
QY 982 cctgcagatggttctgatgttccaaatccagggaactacagttggaggctccaagcaaaa 1041

Db 712 CAAAAAGTGATGCTGAAGAAGTCGCTCTCTCAAGCTAAAAATCGCTGAATTGGAATCAA 771
QY 1042 ggaagtagtattg 1054
Db 772 GTTCATAGACTAG 784

RESULT 15

US-08-246-636-1
Sequence 1, Application US/08246636
Patent No. 5965141

GENERAL INFORMATION:

APPLICANT: Brilles, David E

APPLICANT: Yother, Janet L

APPLICANT: McDaniel, Larry S

APPLICANT: Wu, Hong-yin

TITLE OF INVENTION: EPITOPIC REGIONS OF PNEUMOCOCCAL SURFACE

TITLE OF INVENTION: PROTEIN A

NUMBER OF SEQUENCES: 5

CORRESPONDENCE ADDRESS:

ADDRESSEE: Shoemaker and Mattare, Ltd

STREET: Suite 1203, 2001 Jefferson Davis Highway

CITY: Arlington

STATE: Virginia

COUNTRY: U.S.A.

ZIP: 22202-0286

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/246,636

FILING DATE: 20-MAY-1994

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/656,773

FILING DATE: 15-FEB-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/835,698

FILING DATE: 12-FEB-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/048,896

FILING DATE: 20-APR-1993

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 415-0810

TELEFAX: (703) 415-0813

TELEX: LUKPAT WASHINGTON

INFORMATION FOR SEQ ID NO: 1:

SEQUENCE CHARACTERISTICS:

LENGTH: 2085 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: protein

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORGANISM: Streptococcus pneumoniae

STRAIN: Rxl

IMMEDIATE SOURCE:

CLONE: JY4313

FEATURE:

NAME/KEY: Intron

LOCATION: 1..2085

FEATURE:

NAME/KEY: CDS

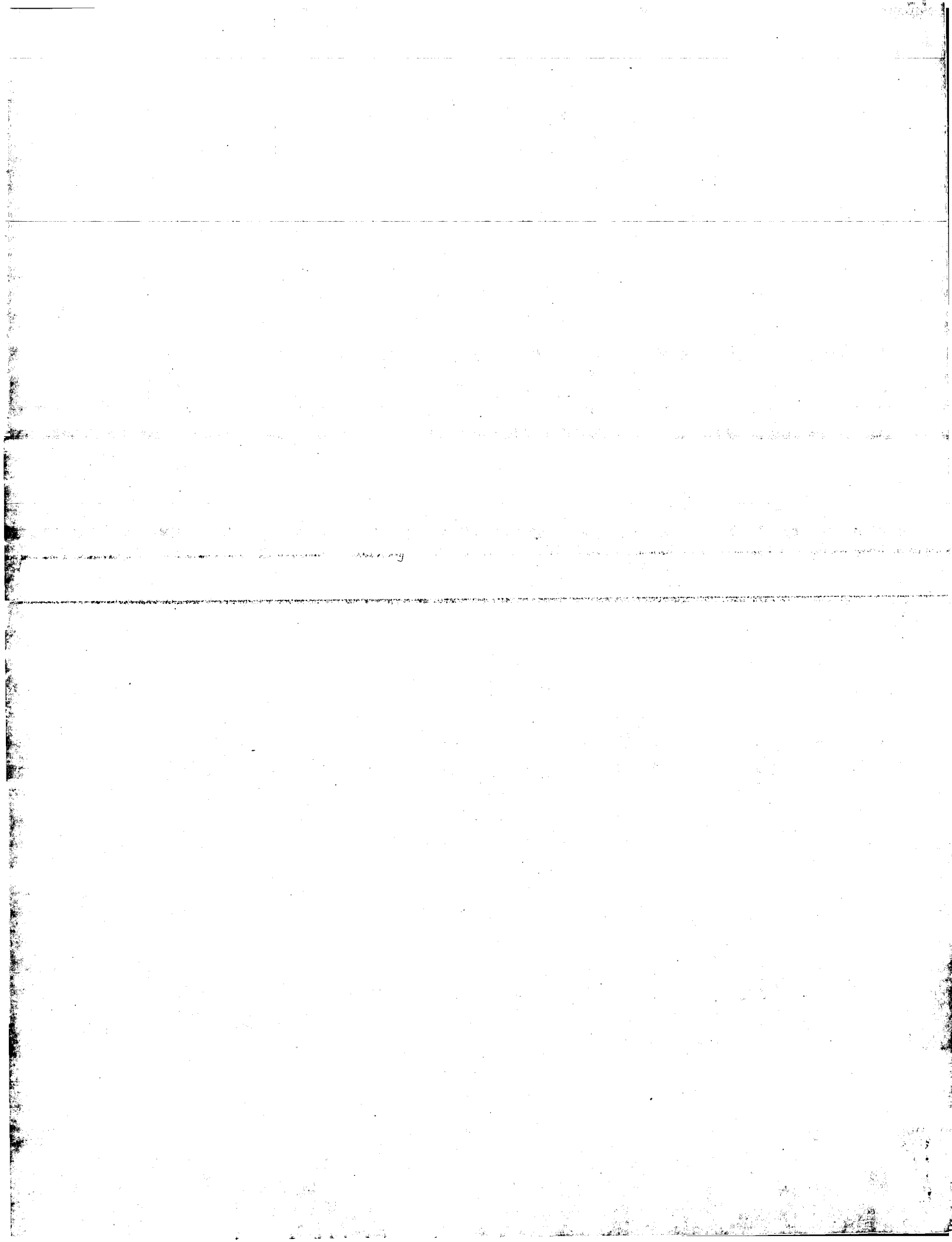
LOCATION: Join(127..1984)

US-08-246-636-1

Query Match 3.4%; Score 50.6; DB 2; Length 2085;
Best Local Similarity 49.8%; Pred. No. 4.3e-05;

	Matches 156;	Conservative 0;	Mismatches 154;	Indels 3;	Gaps 1;
QY 745	atagagaaagatggaatcgattggggatgcataatttgcaggacagaaacgctagtga	804			
Db 472	ACAGACAAAGCCGCAAAAGACGACGACAGATAGATGATGAAAGCTAAGAAACGCGAA	531			
QY 805	gctgtagaaatgctaaatctaatcaacagataaagcaacatagattcagctaaagcagca	864			
Db 532	GAAGAGGCAAAACATAATTTAATCTGTTCGAGCAATGGTAGTTCTGAGCCAGAGCAG	591			
QY 865	atcgctactgctaaagacacaaatagctgaagctcagaaaaagttcccca----ctctcca	921			
Db 592	TTGGCTGAGACTAAGAAAAAATCAGAAAGAGCTAAACAAAAAGCACAGAACTTACTAAA	651			
QY 922	attctcaagagcggaacaaatggtaatcacaggctgagaaagatcttaaaatatcaaa	981			
Db 652	AACTAGAGAGAGCTAAACCAAAATTAGAGAGGCTGAGAAAAAAGCTACTGAAGCCAAA	711			
QY 982	ctgcagatggttctgtatgttccaaatccaggaactcacagttggagggtcccaagcaaca	1041			
Db 712	CAAAAAGTGATGCTGAAGAAGTCGCTCCTCAAGCTAAATCGCTGAATTTGGAAAAATCAA	771			
QY 1042	ggaagtagtatg	1054			
Db 772	GTCATAGACTAG	784			

Search completed: February 7, 2002, 15:53:51
Job time: 345 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run On: February 7, 2002, 21:32:15 ; Search time 7419.31 Seconds
(without alignments)
2129.079 Million cell updates/sec

Title: US-09-391-606-3

Perfect score: 1470

Sequence: 1 atgtgtaactctattgtgcc.....accatcatcatcatcat 1470

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

- 1: em_estfun:*
- 2: em_esthum:*
- 3: em_estin:*
- 4: em_estom:*
- 5: em_estpl:*
- 6: em_estba:*
- 7: em_estro:*
- 8: em_estov:*
- 9: em_hic:*
- 10: gb_est1:*
- 11: gb_est2:*
- 12: gb_hic:*
- 13: gb_gss:*
- 14: em_gss_fun:*
- 15: em_gss_hum:*
- 16: em_gss_inv:*
- 17: em_gss_pln:*
- 18: em_gss_pro:*
- 19: em_gss_rod:*
- 20: em_gss_vrt:*
- 21: em_gss_other:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
C 1	45	3.1	861	13	AZ682538
C 2	45	3.1	906	13	AZ550193
C 3	44.4	3.0	484	10	AA314486
C 4	42.8	2.9	856	13	BH132709
C 5	42.6	2.9	460	10	AA672611
C 6	42	2.9	624	11	BF65301
C 7	41.8	2.8	586	13	AZ632588
C 8	41.8	2.8	850	10	AV406016
C 9	41.6	2.8	762	10	BE661568
C 10	41.4	2.8	842	13	AZ536150
C 11	41.4	2.8	845	13	AZ685553
C 12	41.4	2.8	876	13	AZ691916

13	41.2	2.8	616	13	AZ525704
C 14	41.2	2.8	737	10	AU075947
C 15	41.2	2.8	1000	11	BG291509
C 16	40.8	2.8	841	13	AZ676761
C 17	40.6	2.8	450	11	BG893782
C 18	40.6	2.8	527	13	AQ227560
C 19	40.6	2.8	1042	13	CNS0148K
C 20	40.4	2.7	1101	13	CNS00L4E
C 21	39.8	2.7	915	13	CNS07CXR
C 22	39.8	2.7	938	13	CNS006TJ
C 23	39.6	2.7	422	11	RI5357
C 24	39.6	2.7	502	13	B60448
C 25	39.6	2.7	720	10	BE783140
C 26	39.6	2.7	946	11	BG339626
C 27	39.4	2.7	262	10	BA454826
C 28	39.4	2.7	618	11	RI394077
C 29	39.4	2.7	856	13	AZ667395
C 30	39.2	2.7	298	10	AI218089
C 31	39.2	2.7	384	11	RO3975
C 32	39.2	2.7	695	10	BE217675
C 33	39.2	2.7	855	13	AZ688493
C 34	39.2	2.7	863	13	AZ548883
C 35	39.2	2.7	885	13	AZ693152
C 36	39.2	2.7	938	13	AZ549589
C 37	39	2.7	468	13	AZ522521
C 38	39	2.7	576	11	BG602389
C 39	39	2.7	748	11	BG124392
C 40	39	2.7	884	13	AZ136778
C 41	39	2.7	904	13	AZ692954
C 42	39	2.7	935	13	CNS033D4
C 43	39	2.7	1013	10	AL564621
C 44	38.8	2.6	450	13	FR0025683
C 45	38.8	2.6	738	10	AV405991

ALIGNMENTS

RESULT 1
LOCUS AZ682538/c 861 bp DNA GSS 14-DEC-2000
DEFINITION ENTIALITF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
ACCESSION AZ682538
VERSION AZ682538.1 GI:11819684
KEYWORDS GSS.
SOURCE Entamoeba histolytica.
ORGANISM Entamoeba histolytica.
REFERENCE 1 (bases 1 to 861)
AUTHORS Loftus,B., Van Aken,S. and Fraser,C.
TITLE Determination of clone end sequences from Entamoeba histolytica
JOURNAL HMI:IMSS sheared DNA library
COMMENT Unpublished (2000)
Contact: Brendan J Loftus
Department of Eukaryotic Genomics
The Institute for Genomic Research
9712 Medical Center Dr., Rockville, MD 20850, USA
Tel: 301 838 0200
Fax: 301 838 0208
Email: entaetigr.org

Clones are derived from the Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: shotgun
High quality sequence start: 36
High quality sequence stop: 813.
Location/Qualifiers
1. 861
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"

Geisel, S., Kucaba, T., Lacy, M., Le, M., Martin, J., Morris, M., Schellenberg, K., Steptoe, M., Tan, F., Underwood, K., Moore, B., Theising, B., Wylie, T., Lennon, G., Soares, B., Wilton, R. and Waterston, R.
The WashU-HMI Mouse EST Project
Unpublished (1996)
Contact: Maria M/Mouse EST Project
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
Tel: 314 286 1800
Fax: 314 286 1810
Email: mouseest@wustl.edu
This clone is available royalty-free through LLNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
Seq primer: -28m13 rev2 ET from Amersham
High quality sequence stop: 453.

FEATURES
source
1..460
Location/Qualifiers
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="IMAGE:1054225"
/clone_lib="Soares_mammary_gland_NbMWG"
/sex="male"
/tissue_type="mammary gland"
/dev_stage="4 weeks"
/lab_host="DH10B"
/note="Organ: mammary gland; Vector: pTT73D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA was primed with a Not I - oligo(dT) primer [5', TGTACCAATCTGAAGTGGAGGCGGCCGAATGTTTTTTTTTTTTTTTTTTT T 3']; double-stranded cDNA was ligated to Eco RI adaptors (Pharmacia), digested with Not I and cloned into the Not I and Eco RI sites of the modified pTT73 vector. RNA provided by Dr. Minoru Ko, Wayne State Univ. Library constructed and normalized by Bento Soares and M.Patima Bonaldo."
BASE COUNT 123 a 162 c 100 g 75 t

Query Match 2.9%; Score 42.6; DB 10; Length 460;
Best Local Similarity 53.3%; Pred. No. 1.3;
Matches 90; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
QY 1158 tctgtattcgaagctgcccacagagctgcagacaagctagagcagcgaagccgc 1217
|||||
DB 211 TGCTGTGCTGTGTTGCTGGAGCTGCTAACTGGTGTGCTGGAGCTGCTGCTGC 152
QY 1218 tggagatgacagtgtgtgcagcgtgcagatctcagaagatttagaagcgcctc 1277
|||||
DB 151 AACTGATGAAGCTGCTGCTGCGAGGCGATGCTGCTGAATGGCTGCGCTGTTGA 92
QY 1278 aggtaaagctgggcaacacagcagcctactcaactctttggacagatc 1326
|||||
DB 91 GGACGATGCAACTGCTGCTGAGGGTGTAATGTTGCTGCTGGAACGGAGC 43

RESULT 6
BF865301/c 624 bp mRNA EST 19-JAN-2001
LOCUS 963058F02.y1 C: reinhardtii CC-1690, Stress condition I, normalized
DEFINITION , Lambda Zap II Chlamydomonas reinhardtii cDNA, mRNA sequence.
ACCESSION BF865301
VERSION BF865301.1 GI:12255445
KEYWORDS EST.
ORGANISM Chlamydomonas reinhardtii.
Chlamydomonas reinhardtii.
Eukaryota; Viridiplantae; Chlorophyta; Chlorophyceae; Volvocales;
Chlamydomonadaceae; Chlamydomonas.

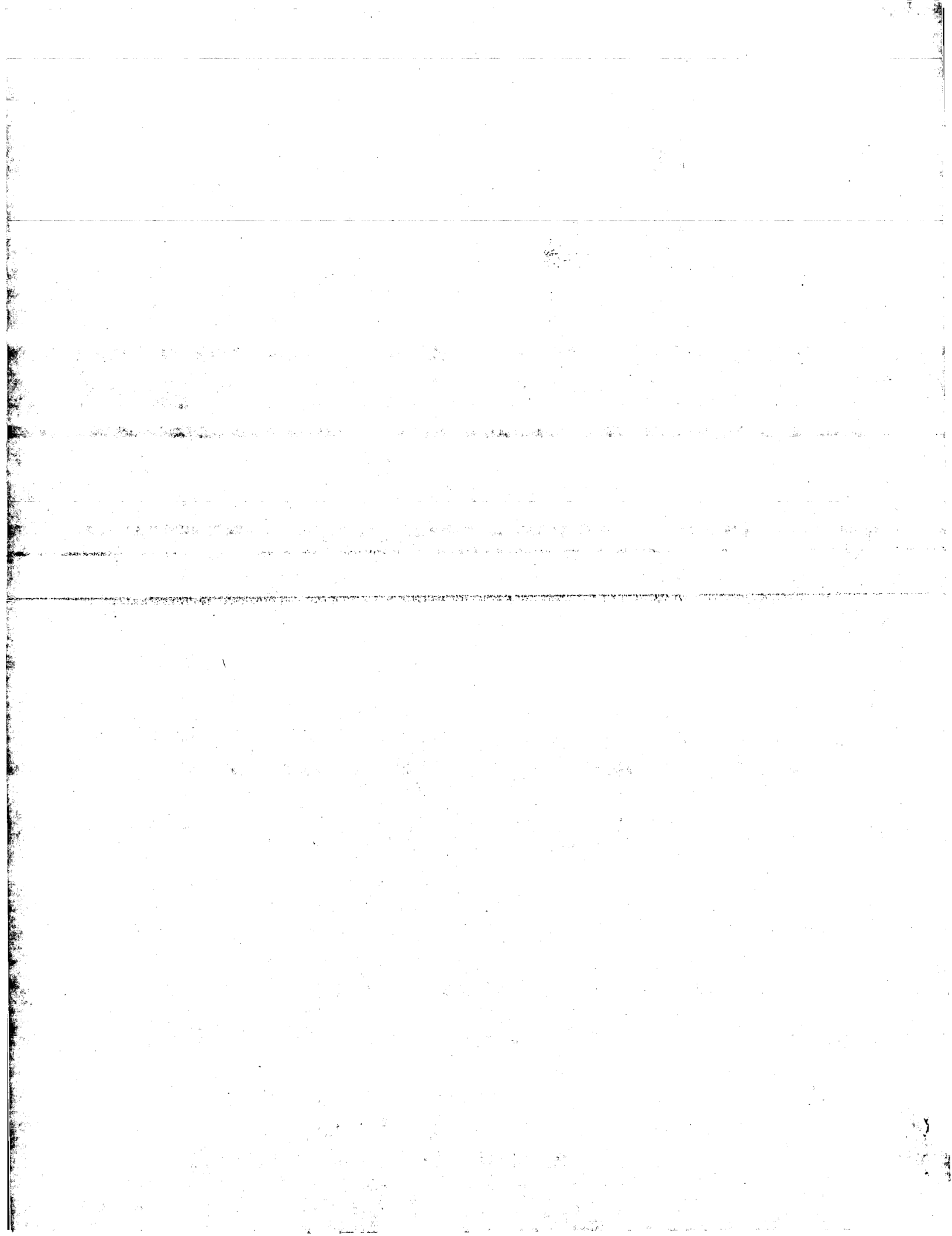
REFERENCE 1 (bases 1 to 624)
Grossman, A., Davies, J., Federspiel, N., Harris, E., Hauser, C., Lefebvre, P., McDermott, J. P., Shrager, J., Silflow, C. and Stern, D. Analyses of the Chlamydomonas reinhardtii Genome: A Model, Unicellular System for Analyzing Gene Function and Regulation in Vascular Plants; project phase 3
Unpublished (2000)
Contact: Charles Hauser
DCMB Box 91000
Duke University
Durham, NC 27708-1000
Tel: 919 613 8159
Fax: 919 613 8177
Email: chauser@duke.edu.

FEATURES
source
1..624
Location/Qualifiers
/organism="Chlamydomonas reinhardtii"
/strain="CC-1690 wild type mt+ 2lgr"
/db_xref="taxon:3055"
/clone_lib="C. reinhardtii CC-1690, Stress condition I, normalized, Lambda Zap II"
/note="Vector: pBluescript II SK-; Site_1: EcoRI; Site_2: XhoI; This library, constructed by John Davies and Jeffrey McDermott, combines cDNAs from CC-1690 cells grown to mid-log phase in TAP-N (30 min, 1hr, 4hr), TAP-S (30 min, 1hr, 4hr), TAP-P (4hr, 12hr, 24hr), NO3 to NH4 (30min, 1hr, 4hr) and NH4 to NO3 (30min, 1hr, 4hr). PolyA mRNA was purified from each sample, pooled and cDNA synthesized. The cDNA was directionally cloned into lambda Zap II (Stratagene) in the EcoRI (5') and XhoI (3') sites. pBluescript II SK- plasmids were excised from the lambda Zap clones by superinfection with ExAssist (Stratagene) phage. The library was normalized using method 4 described in Bonaldo et al (1996) Genome Research 6: 791-806."
BASE COUNT 45 a 138 c 71 g 367 t

Query Match 2.9%; Score 42; DB 11; Length 624;
Best Local Similarity 46.3%; Pred. No. 2.1;
Matches 138; Conservative 0; Mismatches 160; Indels 0; Gaps 0;
QY 684 agggaaacgcctgcaattgtctcaattttgattgacacagatgctacagcacaca 743
|||||
DB 621 AGATAAGCAAGCAATCAATAGAAAAGATAGAAAAGAGAGAGACAGAGAGAGAG 562
QY 744 gatagagaagcagtggaatgcgattggggatgcatatttgcagcagcagcagcgtagt 803
|||||
DB 561 AAAAAAGAAAAAAGAAATAGAAAAAGAAAAAGAGAGAGAGAAAAAAGAAAAAG 502
QY 804 agctgtagaaatgctaaatctataacagctataagcaacatagattcagcctaagcagc 863
|||||
DB 501 AGAAAAAGAAAAACAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 442
QY 864 aatcgctactgctaaagacacaaatagctgaagctcagaaaaagttccccactccaat 923
|||||
DB 441 AACAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 382
QY 924 tcttcagaagcgggacaaatggttaacacagctgagaaagatcttataaaatatccaa 981
|||||
DB 381 AGACAAAAAGAGAGAAAAAGAGAAAAAGAAAAAGAAAAAGAAAAAGAAAAAGAAAA 324

RESULT 7
AZ632588/c 586 bp DNA GSS 13-DEC-2000
LOCUS LM0487L117F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
DEFINITION clone UUGC1M0487L17 F, DNA sequence.
ACCESSION AZ632588
VERSION AZ632588.1 GI:11754778
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus

Search completed: February 7, 2002, 21:32:23
Job time: 20657 sec



XX 09-NOV-2000.
 PD 03-MAY-2000; 2000WO-CA00511.
 PF 03-MAY-1999; 99US-0132270.
 XX 30-JUN-1999; 99US-0141276.
 PR (AVET) AVENTIS PASTEUR LTD.
 XX
 PA Murdin AD, Oomen RP, Wang J, Dunn P;
 XX PI WPI: 2000-687542/67.
 PI P-PSDB; AAY71957.
 DR Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 DR useful for vaccinating against Chlamydia infections -
 XX PT Claim 32; Fig 3; 112pp; English.
 XX
 XX The present sequence is a DNA coding for a fusion protein comprising a
 truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
 CC residues. C. pneumoniae 76 kDa protein is used in the
 CC diagnosis, prevention and treatment of C. pneumoniae
 CC infections (e.g. pneumonia, upper respiratory
 CC tract diseases, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX
 XX Sequence 2238 BP; 689 A; 459 C; 497 G; 593 T; 0 other;
 SQ

Query Match 99.3%; Score 1379.4; DB 21; Length 2238;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1383; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY	1	atgttgaatcctattgttcaggtcctctatagacgaacagacgacacactccgcagat	60
DB	766	atgttgaatcctattgttcaggtcctctatagacgaacagacgacacactccgcagat	825
QY	61	ctttctgctaagattggaggcgagtgacgaataaagatgcgaagctcaagaata	120
DB	826	ctttctgctaagattggaggcgagtgacgaataaagatgcgaagctcaagaata	885
QY	121	gcaggtgcgaagctaaagcttaagaatacgaacgattctgttagcgagatggagcatc	180
DB	886	gcaggtgcgaagctaaagcttaagaatacgaacgattctgttagcgagatggagcatc	945
QY	181	ttcggttctgcagtgaatgctctcatagcttgacagctggaacagtgggtattgcttctagt	240
DB	946	ttcggttctgcagtgaatgctctcatagcttgacagctggaacagtgggtattgcttctagt	1005
QY	241	aacagctcgtcttctactagcagatctgcagacgtggaactcaacgacgacgcgcacct	300
DB	1006	aacagctcgtcttctactagcagatctgcagacgtggaactcaacgacgacgcgcacct	1065
QY	301	acgcctctccaccacgctctgattgattataagactcaagcgcaaacagcttacgatact	360
DB	1066	acgcctctccaccacgctctgattgattataagactcaagcgcaaacagcttacgatact	1125
QY	361	atctttactcaacatcactagctgacatacagctgctttgttgagcctccagatgct	420
DB	1126	atctttactcaacatcactagctgacatacagctgctttgttgagcctccagatgct	1185
QY	421	gtcactaatataaaggatacagcggtcctgattgagggaacacgcaaatgcgtcgagatgg	480
DB	1186	gtcactaatataaaggatacagcggtcctgattgagggaacacgcaaatgcgtcgagatgg	1245
QY	481	gaaactaagaatgccgatgcaattaaagtggcgcaataattacagaattagcgaataat	540
DB	1246	gaaactaagaatgccgatgcaattaaagtggcgcaataattacagaattagcgaataat	1305

pneumonia; upper respiratory tract disease; bronchitis; sinusitis; acute respiratory disease; cough; sore throat; hoarseness; fever; vaccine; immunisation; ds.

Chlamydia pneumoniae.

Key Location/Qualifiers
CDS 101..2056
FT /*tag= a
ET /product= "Chlamydia pneumoniae 76kDa protein"

WO200066739-A2.

09-NOV-2000.

03-MAY-2000; 2000WO-CA00511.

03-MAY-1999; 99US-0132270.

30-JUN-1999; 99US-0141276.

(AVET) AVENTIS PASTEUR LTD.

Murdin AD, Oomen RP, Wang J, Dunn P;

WPI; 2000-687542/67.

P-PSDB; RAY71954.

Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae, useful for vaccinating against Chlamydia infections -

Claim 2a; Fig 1; 112pp: English.

The present sequence is a DNA coding for Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa protein is used in the diagnosis, prevention and treatment of C. pneumoniae infections (e.g. pneumonia, upper respiratory tract disease, bronchitis, sinusitis and acute respiratory disease such as cough, sore throat, hoarseness, fever; and abnormal chest sounds on auscultation). C. pneumoniae sequence is also used as vaccines for immunising humans against diseases caused by C. pneumoniae.

Sequence 2156 BP; 672 A; 461 C; 471 G; 552 T; 0 other;

Query Match 97.9%; Score 1359.4; DB 21; Length 2156;
Best Local Similarity 99.5%; Pred. No. 0;
Matches 1374; Conservative 0; Mismatches 6; Indels 1; Gaps 1;
Qy 1 atggttaactctattggtccagggtcctatagacgaaacagacgacacccctccgcagat 60
Db 101 atggttaactctattggtccagggtcctatagacgaaacagacgacacccctccgcagat 160
Qy 61 cttctctcgaagattgagcgagtcgacaaataagtcggaagtcggaagtcggaagata 120
Db 161 cttctctcgaagattgagcgagtcgacaaataagtcggaagtcggaagtcggaagata 220
Qy 121 gcaggtcggaagctaaagcctaaagaaatcctaaagcagattctgtagacgagtcgagc 180
Db 221 gcaggtcggaagctaaagcctaaagaaatcctaaagcagattctgtagacgagtcgagc 280
Qy 181 ttgggttcgagtgaaatgctctcatagtcgagtcgagtcgagtcgagtcgagtcgagtc 240
Db 281 ttgggttcgagtgaaatgctctcatagtcgagtcgagtcgagtcgagtcgagtcgagtc 340
Qy 241 aacagctcgtctctactagcagatctgacgagtcgagtcgagtcgagtcgagtcgagtc 300
Db 341 aacagctcgtctctactagcagatctgacgagtcgagtcgagtcgagtcgagtcgagtc 400
Qy 301 acgctctccaccacgctctgattgattatagactcaagcgaacacgcttacgatact 360
Db 401 acgctctccaccacgctctgattgattatagactcaagcgaacacgcttacgatact 460

Qy 361 atctttactcaacatcactagtcacatacagcgtctgttggtagcctccagatgct 420
Db 461 atctttactcaacatcactagtcacatacagcgtctgttggtagcctccagatgct 520
Qy 421 gtactataataaggatatacagcgtctactgtatgaggaaacgcgaatcgctcgagtg 480
Db 521 gtactataataaggatatacagcgtctactgtatgaggaaacgcgaatcgctcgagtg 580
Qy 481 gaaactaagaatccgatgcaatgaaatgagcgcgaatgaaatgaaatgaaatgaaatg 540
Db 581 gaaactaagaatccgatgcaatgaaatgagcgcgaatgaaatgaaatgaaatgaaatg 640
Qy 541 gcttcgataacaaagcgtattctgactctttagttaaactgactctcttcgactctta 600
Db 641 gcttcgataacaaagcgtattctgactctttagttaaactgactctcttcgactctta 700
Qy 601 cagactgctcttctccatctgtagcaacaatacaaaagcgtgagctctttaaagag 660
Db 701 cagactgctcttctccatctgtagcaacaatacaaaagcgtgagctctttaaagag 760
Qy 661 atgcaagataacccagtagtcccgaggaacgcgcgtcaattgctcaattcttagttg 720
Db 761 atgcaagataacccagtagtcccgaggaacgcgcgtcaattgctcaattcttagttg 820
Qy 721 cagcagatgctacagcagacagatagagaagatggaatgcgattgggagtcacat 780
Db 821 cagcagatgctacagcagacagatagagaagatggaatgcgattgggagtcacat 880
Qy 781 ttgcagagcagacgctgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 840
Db 881 ttgcagagcagacgctgagtcgagtcgagtcgagtcgagtcgagtcgagtcgagtcg 940
Qy 841 aacatagattcagctaaagcagcagcagcagcagcagcagcagcagcagcagcagc 900
Db 941 aacatagattcagctaaagcagcagcagcagcagcagcagcagcagcagcagcagc 1000
Qy 901 aaaaagtctcccgactctcccaattctcaagaagcgaacaaatggttaacagcgtgag 960
Db 1001 aaaaagtctcccgactctcccaattctcaagaagcgaacaaatggttaacagcgtgag 1060
Qy 961 aaagattctaaataatcaaacctcagatggttctgattgttcccaatccaggaactaca 1020
Db 1061 aaagattctaaataatcaaacctcagatggttctgattgttcccaatccaggaactaca 1120
Qy 1021 gttgaggtctcaagcaacaggaagtagtattgttagtattgttccatgctgtta 1080
Db 1121 gttgaggtctcaagcaacaggaagtagtattgttagtattgttccatgctgtta 1180
Qy 1081 gatgtagtgaataatgagacgcgttccattttgattgctgggttctcagatgattcac 1140
Db 1181 gatgtagtgaataatgagacgcgttccattttgattgctgggttctcagatgattcac 1240
Qy 1141 atgttcaatcaggaataatcctgattctcaagcgtcccaacagagcgtcgacacacgct 1200
Db 1241 atgttcaatcaggaataatcctgattctcaagcgtcccaacagagcgtcgacacacgct 1300
Qy 1201 agagcagcgaagcgcgtggagtagacagtcgtctgctgagcgtcgagatgctcagaaa 1260
Db 1301 agagcagcgaagcgcgtggagtagacagtcgtctgctgagcgtcgagatgctcagaaa 1360
Qy 1261 gctttagaagcgcgtctcaggttaaaagtcgggcaacaacaggggcataactcaattcttggga 1320
Db 1361 gctttagaagcgcgtctcaggttaaaagtcgggcaacaacaggggcataactcaattcttggga 1420
Qy 1321 caagcgtctctgctgctgtgtgtgagcagcagag-tcctcccgctgcagcagattctatg 1379
Db 1421 caagcgtctctgctgctgtgtgtgtgagcagcagag-tcctcccgctgcagcagattctatg 1480
Qy 1380 g 1380
Db 1481 g 1481

RESULT 3
AAx91990/c
ID AAX91990 standard; DNA; 1230025 BP.
XX AC
AC AAX91990;
DT 13-SEP-1999 (first entry)
XX DE Nucleotide sequence of the complete genome of Chlamydia pneumoniae.
XX DE Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope; ss.
XX OS Chlamydia pneumoniae.
XX PN WO927105-A2.
XX PD 03-JUN-1999.
XX PF 20-NOV-1998; 98WO-IB01890.
XX PR 04-NOV-1998; 98US-0107078.
XX PR 21-NOV-1997; 97FR-0014673.
XX PA (GEST) GENSET.
PI Griffiths R;
XX WPT: 1999-357842/30.
XX PT Genome sequence of Chlamydia pneumoniae
PS Claim 1; Page 291-611; 1912pp; English.
CC The present sequence represents the complete genome of Chlamydia
CC pneumoniae, and encodes proteins AAY34584-Y35879. C. pneumoniae causes
CC respiratory disease such as pneumonia and bronchitis and is thought
CC to be a contributing factor in heart disease, sarcoidosis, sinusitis,
CC purulent otitis media, erythema nodosum or pharyngitis. The polypeptides
CC encoded by the open reading frames of the C. pneumoniae genome (see
CC AAY34584-Y35879) can be used in immunogenic compositions as vaccines.
CC Vectors containing C. pneumoniae nucleotides sequences can also be
CC used as immunogenic compositions, especially where the vector directs
CC the expression of a neutralising epitope of C. pneumoniae.
XX SQ Sequence 1230025 BP; 367213 A; 249833 C; 249013 G; 363589 T; 377 other;
Query Match 97.1%; Score 1348.4; DB 20; Length 1230025;
Best Local Similarity 99.4%; Pred. No. 0;
Matches 1374; Conservative 0; Mismatches 6; Indels 2; Gaps 2

QY	1	atggttaatcctatttggtccaggttcctatagacgaacagaacgcacacctcccagat	60
Db	828598	ATGTTTAACTCCTATTGGTCAGGTCCATAGCGAACAACAGACGCACACTCCCAGAT	828599
QY	61	ccttctgctcaaggattggcgagtcgagcaataaagagtgcggaagctcaaagaata	120
Db	828538	CYTTCGCTCAAGGATTGAGCGGAGTCGAGCAAATAAGAGTGCGGAAGCTCAAGAATA	828479
QY	121	gcaggtgcggaagctaagcctaaagaactaaagaccgattctgtagcgatggagcatc	180
Db	828478	GCAGGTGCGGAAGCTTAAGCCTTAAGAATCTAAGACCAGATTCTGTAGAGCATGGAGCATC	828419
QY	181	tigcgtctcgagtgaattgctctcatgagtcgtgcagataaagctgggtatttgtttctag	240
Db	828418	TTCGGTCTCGAGTGAATGCTCTCATGAGTCGSCAGATAAGCTGGGTATTGCTTCTAGT	828359
QY	241	aacagctcgtctcttactagcagatctgcagagctggaactcaacgacgcacccacct	300
Db	828358	AACAGCTCGTCTTCTACTAGCAGATCTGCAGCTGTGACTCAACGACGCCGAGCACCT	828299

205 atgagctctggcagataaagctgggcttctgtcttagtaacagctgctctctactacga
 1 atgagctctggcagataaagctgggcttctgtcttagtaacagctgctctctactacga
 265 tctgcagagctggactcaacgacagcgacgcctctctctccacccagctctgag
 61 tctgcagagctggactcaacgacagcgacgcctctctctccacccagctctgag

121	gattataagactcaagcgcaaacagcttacgatactatctttacctcaacatacactagc
Db	
385	gacatacaggctcttggtagctccaggatgcttcactaataataaagatcacagc
Qy	
181	gacatacaggctcttggtagctccaggacgctgcactaataataaagatcacagc
Db	
445	gctactgatgaggaaaccgcaatcgcctgcggagtgggaaactaagaatgccgatgcata
Qy	

Qy	505	aaagttggcgcgaattacagaaattagcgaataatgcttcggataaccaacgcgattctt
Db	301	aaagttggcgcgaattacagaaattagcgaataatgcttcggataaccaacgcgattctt

Qy 565 gactcttttaggtaaactgactcttccttlogacctctttaacagactgctcttcttccaatctg

Qy	625	gcaacaataa	caaacagctg	gagcttctt	taagagatg	caagataaac	cccgtagtcc
Db	421	gcaacaataa	caaacagctg	gagcttctt	taagagatg	caagataaac	cccgtagtcc
Qy	685	gggaaacgc	ctgaaattg	ctcaatttt	tagttgta	taagacagat	gctcacgcgaca

Db 481 gggaaaacgcctgcaattgctcaatctttagttgattcagacagatgctacagcgacac

Qy 745 atagagaaagatggaaaatgcgattggggatgcataatttcagggacagaaacgcctagtg

Db 541 atagagaaagatggaatgcgattagggatgcataatttgcaggacagaaacgctagtgc

Qy 805 gctgtagaaaatgctaaatctaataacagtataagcaacatagattcagctaaagcagc

Db 601 gctgtagaaaatgctaaatctaataacagttataagcaacatagattcagctaaagcag

Qy 865 atccgtactgctaagacacaaatagctgaagctcagaaaaagtccccgactctccaa

661 atcgctactgctaagacacacaaatagctgaagctcagaaaaaagtccccgactctccaa

925 cttcaagaaacggaacaaatggttaatacacagcctgagaaatcttaaaaatatcaaac

721 ctccaagaaacggaacaaatggtaatcacaggctgagaaagatcttaaaaaatatcaaac

985 gcagatggttctgatgttccaataccaggaaactacagttagggtccaagcaacaag

[illegible]

pB agtgaagctgatacgtatgcgtgtttcccatgctgtttatatgatgctgaaacattgaagaacgg
qy 1045 agtagtatggtagtaacctgtcccccaagtcaccagaagagctcgaaaaaacgagacac

QY	1105	tccatttggatgctgggttttcgcagatgattcacatggttcacaacggaataatctctg
Db	901	tccatttggatgctgggttttcgcagatgattcacatggttcacaacggaataatctctg
QY	1105	tctcaagctgcccaacagggaactcgagcagcacaagctagacgacgaaagccgtgcgag

Db	961	tctcaagctgcccacacaggagctcgacagacacaagctagagcagcgaaagccgctgagat	1020
Qy	1225	gacagtgtctgtcgacgctgtggcagatgctcagaaagcttttagaagcggtcttaggtaaa	1284
Db	1021	gacagtgtctgtcgacgctgtggcagatgctcagaaagcttttagaagcggtcttaggtaaa	1080
Qy	1285	gctggggcaacaacaggggcatactcaatgcttttggacagatcgcttctgtgctgtttgtg	1344
Db	1081	gctfgggcaacaacaggggcatactcaatgcttttagacagatcgcttctgtgctgtttgtg	1140
Qy	1345	agcgcaggag-tectccgctgcagcaagttctatgg	1380
Db	1141	aqcgaagagtttcttcccgctacacaaagttctatg	1177

RESULT 6
AAV16207
ID AAV16207 standard; DNA; 150 BP.
XX
XX AC AAV16207;
XX
XX DT 28-MAY-1998 (first entry)
XX
XX DE Part of the gene encoding the 76 kDa cystein rich OMP.
XX
XX KW Cysteine rich outer-membrane protein; OMP; 60 kDa OMP; 76 kDa OMP;
KW hybridisation; amplification; assay; detection; ds.
XX
XX OS Chlamydia pneumoniae.
OS
XX WO9746709-A2.
PN
XX
XX PD 11-DEC-1997.
XX
XX PF 04-JUN-1997; 97WO-US09673.
XX
XX PR 06-JUN-1996; 96US-0659473.
XX
XX (ABBO) ABBOTT LAB.
PA
XX
XX PI Cerney MB;
XX
XX DR WPI; 1998-042215/04.
XX
XX PT Chlamydia pneumoniae derived oligonucleotides - used as primers or
PT probes for specific and sensitive detection
XX
XX PS Disclosure; Page 24; 28pp; English.
XX
XX CC The present sequence represents part of the gene encoding the 76 kDa
CC cysteine rich outer-membrane protein (OMP) of Chlamydia pneumoniae.
CC Novel oligonucleotides AAV16195-206 can function either as PCR primers
CC or probes. They are used to detect C. pneumoniae in a sample.
CC Oligonucleotides V161202-04 are specific for the present sequence.
CC The oligonucleotides can be used in a hybridisation or amplification
CC based assay for the detection of C. pneumoniae in a test sample.
XX
XX SQ Sequence 150 BP; 43 A; 34 C; 40 G; 33 T; 0 other;

Db.					
	121 taagaatccgatgcagtgtaagtggcgc	150			
	RESULT 7				
	AAF58252/C				
ID	AAF58252 standard; DNA; 936 BP.				
XX	AAAF58252;				
XX	24-APR-2001: (first entry)				
DE	Oligonucleotide D1835.				
KW	Electron-transfer group; ETM; mismatch;				
KW	gene expression; ss.				
OS	Synthetic.				
PN	WO200107665-A2.				
XX	01-FEB-2001.				
PF	26-JUL-2000; 2000WO-US20476.				
PR	26-JUL-1999; 99US-0145695.				
PR	17-MAR-2000; 2000US-0190259.				
XX	(CLIN-) CLINICAL MICRO SENSORS INC.				
PA	Umek RM;				
PI	WPI; 2001-159728/16.				
DR	Nucleic acids containing electron-trans				
PT	hybridization assays, e.g. for genotypi				
PT	a single surface				
XX	Example 6; Page 127; 159pp; English.				
PS	The present invention relates to a compo				
XX	acids each containing an electron-trans				
CC	different redox potentials. The inventi				
CC	detection of nucleic acids, especially				
CC	and single-nucleotide polymorphisms, e.				
CC	monitoring gene expression.				
XX	Sequence 936 BP; 4 A; 139 C; 10 G; 7 T;				
SO					

PD 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-US20476.
 XX 26-JUL-1999; 99US-0145695.
 XX 17-MAR-2000; 2000US-0190259.
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 XX UmeK RM;
 XX WPI; 2001-159728/16.
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX Example 6; Page 127; 159pp; English.
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX Sequence 936 BP; 5 A; 142 C; 7 G; 6 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 936;
 Best Local Similarity 1.2%; Pred. No. 7.4e-14;
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatccgatgcaattaaagtgtgcgcgcaattacagaattagcgaatag 541
 DB 750 WW 691
 QY 542 ctccggataaccaagcattctgactcttttaggttaaactgactcttccttccttac 601
 DB 690 GWW 631
 QY 602 agactgctcttcccaatctgtagcaacaataacaaagcagctgacttcttaagaga 661
 DB 630 WW 571
 QY 662 tgcaagataaccagtagtcccgaggaaacgcctgcaattgctcaattcttagttgac 721
 DB 570 WW 511
 QY 722 agacagatgctacagcgacacagatagagaagatggaatgcgattgggagatgcatatt 781
 DB 510 WW 451
 QY 782 ttgcagagacagcagtagtgagctgtagaaatgctaaatcttaataacagataagca 841
 DB 450 WW 391
 QY 842 acatagattcagctaaacagcagcaatcgctactgtaagacacaaatagctgagctcaga 901
 DB 390 WW 331
 QY 902 aaaagtcccgactctcccaattcttcaagaagcgaacaaatgtaatacagcgctgaga 961
 DB 330 WW 271
 QY 962 aagattcttaaaataatcaaacctgcagatggttctgattgttcccaatccaggaactacag 1021
 DB 270 WW 211
 QY 1022 ttggaggctccaagcaacaaggaagtagtattgtagttattgctgttccatgctgttag 1081
 DB 210 WW 151

QY 1082 atgatgtgaaatgagaccgcttcatttgatgtctgggttctgcagatgattcaca 1141
 DB 150 WW 91
 QY 1142 tgttcaatacggaaaatcctgattctcaa 1170
 DB 90 WW 62

RESULT 10
 AAF58259/C
 ID AAF58259 standard; DNA; 936 BP.
 XX AAF58259;
 XX 24-APR-2001 (first entry)
 XX Oligonucleotide D2004.
 XX Electron-transfer group; ETM; mismatch; genotyping;
 KW gene expression; ss.
 XX Synthetic.
 XX WO200107665-A2.
 XX 01-FEB-2001.
 XX 26-JUL-2000; 2000WO-US20476.
 XX 26-JUL-1999; 99US-0145695.
 XX 17-MAR-2000; 2000US-0190259.
 XX (CLIN-) CLINICAL MICRO SENSORS INC.
 XX UmeK RM;
 XX WPI; 2001-159728/16.
 XX Nucleic acids containing electron-transfer group, useful as labels in
 PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
 PT a single surface
 XX Example 6; Page 128; 159pp; English.
 XX The present invention relates to a composition comprising two nucleic
 CC acids each containing an electron-transfer group (ETM) having
 CC different redox potentials. The invention is used for electronic
 CC detection of nucleic acids, especially of substitutions (mismatches)
 CC and single-nucleotide polymorphisms, e.g. for genotyping,
 CC monitoring gene expression.
 XX Sequence 936 BP; 6 A; 138 C; 8 G; 8 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 936;
 Best Local Similarity 1.2%; Pred. No. 7.4e-14;
 Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

QY 482 aaactaagaatccgatgcaattaaagtgtgcgcgcaattacagaattagcgaatag 541
 DB 750 WW 691
 QY 542 ctccggataaccaagcattctgactcttttaggttaaactgactcttccttccttac 601
 DB 690 GWW 631
 QY 602 agactgctcttcccaatctgtagcaacaataacaaagcagctgacttcttaagaga 661
 DB 630 WW 571
 QY 662 tgcaagataaccagtagtcccgaggaaacgcctgcaattgctcaattcttagttgac 721


```

xx OS Synthetic.
xx PN WO200107665-A2.
xx PD 01-FEB-2001.
xx PF 26-JUL-2000; 2000WO-US20476.
xx PR 26-JUL-1999; 99US-0145695.
xx PR 17-MAR-2000; 2000US-0190259.
xx PA (CLIN-) CLINICAL MICRO SENSORS INC.
xx PI Umek RM;
xx DR WPI; 2001-159728/16.
xx PT Nucleic acids containing electron-transfer group, useful as labels in
xx PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
xx PT a single surface
xx PS Example 6; Page 127; 159pp; English.
xx CC The present invention relates to a composition comprising two nucleic
xx CC acids each containing an electron-transfer group (ETM) having
xx CC different redox potentials. The invention is used for electronic
xx CC detection of nucleic acids, especially of substitutions (mismatches)
xx CC and single-nucleotide polymorphisms, e.g. for genotyping,
xx CC monitoring gene expression.
xx SQ Sequence 938 BP; 4 A; 144 C; 9 G; 5 T; 776 other;

Query Match 6.0%; Score 83; DB 22; Length 938;
Best Local Similarity 1.2%; Pred. No. 7.4e-14;
Matches 8; Conservative 403; Mismatches 278; Indels 0; Gaps 0;

Qy 482 aaactaagaatgcgagatcaattaaagtggcgcaaatcacagaattagcgaataatg 541
Db 750 www. .... 691
Qy 542 ctccgataacacagcagattctgactcttagtgtaaaactgactcttcgcactcttac 601
Db 690 Gwww. .... 631
Qy 602 agactgctcttccaatctgtagcaacaataacaaagcagctgagcttcttaaaagaga 661
Db 630 www. .... 571
Qy 662 tgcagataacccagtagtccagggaacgcctgcaattgctcaattcttagtgatc 721
Db 570 www. .... 511
Qy 722 agacagatgctacagcagacagatagagaagaatggaatgcgattgggagatcatatt 781
Db 510 www. .... 451
Qy 782 ttgcaggacagacgctagtgagctgtagaaaatgctaaaatctaaataacagataagca 841
Db 450 www. .... 391
Qy 842 acatagattcagctaaacagcagcaatcgctactgctctgctaaacacaaatagctgaagctcaga 901
Db 390 www. .... 331
Qy 902 aaaagtcccgactctccattctcaaggaagcgggaacaaatggaataacagcgtcaga 961
Db 330 www. .... 271
Qy 962 aagattctaaataatcaaccctgcagatggttctgtgttccaaatccaggaaactacag 1021
Db 270 www. .... 211

```

```

Qy 1022 ttgaggctccaagcaagaaagtagtattgtagtattcgtttccatgctgttag 1081
Db 210 www. .... 151
Qy 1082 atgatgctgaaatgagaccgcttcatttgatgctgggttctgcagatgattcaca 1141
Db 150 www. .... 91
Qy 1142 tgttcaatacggaaaaatcctgattctcaa 1170
Db 90 www. .... 62

RESULT 13
AAF58252
ID AAF58252 standard; DNA; 936 BP.
XX AC AAF58252;
XX DT 24-APR-2001 (first entry)
XX DE Oligonucleotide D1835.
XX KW Electron-transfer group; ETM; mismatch; genotyping;
XX KW gene expression; ss.
XX OS Synthetic.
XX PN WO200107665-A2.
XX PD 01-FEB-2001.
XX PF 26-JUL-2000; 2000WO-US20476.
XX PR 26-JUL-1999; 99US-0145695.
XX PR 17-MAR-2000; 2000US-0190259.
XX PA (CLIN-) CLINICAL MICRO SENSORS INC.
XX PI Umek RM;
XX DR WPI; 2001-159728/16.
XX PT Nucleic acids containing electron-transfer group, useful as labels in
XX PT hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX PT a single surface
XX PS Example 6; Page 127; 159pp; English.
XX CC The present invention relates to a composition comprising two nucleic
XX CC acids each containing an electron-transfer group (ETM) having
XX CC different redox potentials. The invention is used for electronic
XX CC detection of nucleic acids, especially of substitutions (mismatches)
XX CC and single-nucleotide polymorphisms, e.g. for genotyping,
XX CC monitoring gene expression.
XX SQ Sequence 936 BP; 4 A; 139 C; 10 G; 7 T; 776 other;

Query Match 5.8%; Score 80.6; DB 22; Length 936;
Best Local Similarity 1.2%; Pred. No. 3.7e-13;
Matches 8; Conservative 402; Mismatches 281; Indels 0; Gaps 0;

Qy 483 aaactaagaatgccatgcaattaaagtggcgcaaatcacagaattagcgaataatgc 542
Db 1 www. .... 60
Qy 543 ttccgataaccaaagcagattctttagcttttaggtaaactgactcttcgcactcttaca 602
Db 61 www. .... 120
Qy 603 gactgctcttccaaatctgtagcaacaataacaaagcagctgagcttcttaagagat 662

```

```

Db 121 ..... 180
Qy 663 gcaagataaccagtagtccacgaggaacacccctgcaattgctcaattttagtgatca 722
Db 181 ..... 240
Qy 723 gacagatgctacagcgacacagatagagaaagatggaatgctggtgggagctcatatt 782
Db 241 ..... 300
Qy 783 tgcaggacagacgctagtgagctgtagaaatgctaaatctataacagatataagcaa 842
Db 301 ..... 360
Qy 843 catagattcagctaaagcagcaatgctactgctaaagacacacaaatagctgaagctcagaa 902
Db 361 ..... 420
Qy 903 aaagtcccccactctccattctcaagaagcggaacaaatggttaacacagctgagaa 962
Db 421 ..... 480
Qy 963 agatcttaaaatatacaacccctgagctgttctgatttccaaatccaggaactacagt 1022
Db 481 ..... 540
Qy 1023 tggagctcccaagcacaagaagtagtattggttagtattcgtttccatgctgttaga 1082
Db 541 ..... 600
Qy 1083 tgatgctgaaatgagacgctctccatttctgattgctggtttcgtcagatgattcacat 1142
Db 601 ..... 660
Qy 1143 gtccaatacggaaatccctgatttccagct 1173
Db 661 ..... 691

```

```

RESULT 14
AAF58254
ID AAF58254 standard; DNA; 936 BP.
AC AAF58254;
DT 24-APR-2001 (first entry)
DE Oligonucleotide D1875.
KW Electron-transfer group; ETM; mismatch; genotyping;
KW gene expression; ss.
OS Synthetic.
XX WO200107665-A2.
XX 01-FEB-2001.
XX 26-JUL-2000; 2000WO-US20476.
XX 26-JUL-1999; 99US-0145695.
XX 17-MAR-2000; 2000US-0190259.
XX (CLIN-) CLINICAL MICRO SENSORS INC.
XX Umek RM;
XX WPI; 2001-159728/16.
XX Nucleic acids containing electron-transfer group, useful as labels in
XX hybridization assays, e.g. for genotyping, allowing repeat analyses on
XX a single surface

```

```

XX Example 6; Page 127; 159pp; English.
XX The present invention relates to a composition comprising two nucleic
XX acids each containing an electron-transfer group (ETM) having
XX different redox potentials. The invention is used for electronic
XX detection of nucleic acids, especially of substitutions (mismatches)
XX and single-nucleotide polymorphisms, e.g. for genotyping,
XX monitoring gene expression.
XX Sequence 936 BP; 4 A; 144 C; 7 G; 5 T; 776 other;
SQ
Query Match 5.8%; Score 80.6; DB 22; Length 936;
Best Local Similarity 1.2%; Pred. No. 3.7e-13;
Matches 8; Conservative 402; Mismatches 281; Indels 0; Gaps 0;
Qy 483 aactaagaatgccgatgcaattaaagtggcgcgcaaatattacagaatagcgaatgc 542
Db 1 ..... 60
Qy 543 ttccgataaacaagcagatttctgactctttaggttaaaactgacttcttcgacctttaca 602
Db 61 ..... 120
Qy 603 gactgtctctccaaatctgtagcaacaatacaaacagcagctgagcttcttaagagat 662
Db 121 ..... 180
Qy 663 gcaagataaccagtagtccacgaggaacacccctgcaattgctcaattttagtgatca 722
Db 181 ..... 240
Qy 723 gacagatgctacagcgacacagatagagaaagatggaatgctggtgggagctcatatt 782
Db 241 ..... 300
Qy 783 tgcaggacagacgctagtgagctgtagaaatgctaaatctataacagatataagcaa 842
Db 301 ..... 360
Qy 843 catagattcagctaaagcagcaatgctactgctaaagacacacaaatagctgaagctcagaa 902
Db 361 ..... 420
Qy 903 aaagtcccccactctccattctcaagaagcggaacaaatggttaacacagctgagaa 962
Db 421 ..... 480
Qy 963 agatcttaaaatatacaacccctgagctgttctgatttccaaatccaggaactacagt 1022
Db 481 ..... 540
Qy 1023 tggagctcccaagcacaagaagtagtattggttagtattcgtttccatgctgttaga 1082
Db 541 ..... 600
Qy 1083 tgatgctgaaatgagacgctctccatttctgattgctggtttcgtcagatgattcacat 1142
Db 601 ..... 660
Qy 1143 gtccaatacggaaatccctgatttccagct 1173
Db 661 ..... 691

```

```

RESULT 15
AAF58257
ID AAF58257 standard; DNA; 936 BP.
XX AAF58257;
XX AAF58257;
DT 24-APR-2001 (first entry)

```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 15:53:51 ; Search time 243.49 Seconds
(without alignments)
1291.954 Million cell updates/sec

Title: us-09-391-606-4

Perfect score: 1389

Sequence: 1 atggttaactctattgttc.....aagttatgatccgagct 1389

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 351203 seqs, 113238999 residues

Total number of hits satisfying chosen parameters: 702406

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_NA.*

- 1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
- 2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
- 3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
- 4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
- 5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq.*
- 6: /cgn2_6/ptodata/2/ina/backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	146.8	10.6	150	4	US-08-659-473-8
2	54.4	3.9	7218	1	US-08-232-463-14
3	50.6	3.6	2085	1	US-08-072-070-1
4	50.6	3.6	2085	1	US-08-465-746-1
5	50.6	3.6	2085	1	US-08-214-164-1
6	50.6	3.6	2085	1	US-08-469-434-1
7	50.6	3.6	2085	1	US-08-214-222-1
8	50.6	3.6	2085	2	US-08-467-852A-1
9	50.6	3.6	2085	2	US-08-468-718-1
10	50.6	3.6	2085	2	US-08-246-636-1
11	50.6	3.6	2085	2	US-08-319-795-1
12	50.6	3.6	2085	2	US-08-468-985-1
13	50.6	3.6	2085	2	US-08-312-949-1
14	50.6	3.6	2086	3	US-08-446-201-2
15	50.6	3.6	2086	3	US-08-119-125A-2
16	43.6	3.1	6744	1	US-08-961-083-33
17	39.6	2.9	640	3	US-08-072-610-1
18	39.4	2.8	3337	1	US-08-719-822B-1
19	39.4	2.8	3337	2	US-09-092-458-1
20	39.4	2.8	3337	1	US-08-186-222-1
21	37.4	2.7	1920	1	US-08-973-462-2
22	37.4	2.7	5361	4	US-09-029-213B-7
23	37.4	2.7	5820	4	US-08-973-462-1
24	37.4	2.7	6152	4	US-08-973-462-1
25	36.4	2.6	5738	3	US-08-409-995-3
26	36.4	2.6	5738	3	US-08-685-467-3
27	36.4	2.6	7291	4	US-08-913-942-3

28	35.6	2.6	2943	2	US-08-788-892-1	Sequence 1, Appli
29	35	2.5	697	6	5171843-10	Patent No. 5171843
30	35	2.5	1137	6	5171843-8	Patent No. 5171843
31	35	2.5	1855	3	US-08-961-083-71	Sequence 71, Appl
32	34.8	2.5	5177	6	5352450-1	Patent No. 5352450
33	34.6	2.5	414	2	US-08-630-822A-63	Sequence 63, Appl
34	34.6	2.5	414	2	US-09-005-069-63	Sequence 63, Appl
35	34.6	2.5	650	4	US-08-998-416-190	Sequence 190, Appl
36	34.6	2.5	708	4	US-09-285-576-1	Sequence 1, Appli
37	34.6	2.5	716	4	US-08-998-416-1032	Sequence 1032, Ap
38	34.6	2.5	835	4	US-08-998-416-533	Sequence 533, App
39	34.6	2.5	836	4	US-08-998-416-286	Sequence 286, App
40	34	2.4	168	1	US-08-469-802B-4	Sequence 4, Appli
41	34	2.4	168	2	US-08-267-803B-4	Sequence 4, Appli
42	34	2.4	171	1	US-08-469-802B-5	Sequence 5, Appli
43	34	2.4	171	2	US-08-267-803B-5	Sequence 5, Appli
44	34	2.4	195	1	US-08-469-802B-2	Sequence 2, Appli
45	34	2.4	195	2	US-08-267-803B-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1
US-08-659-473-8
; Sequence 8, Application US/08659473
; Patent No. 6210876
; GENERAL INFORMATION:
; APPLICANT: M. B. Cerney
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES FOR DETECTING
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Abbott Laboratories
; STREET: 100 Abbott Park Road
; CITY: Abbott Park
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60064-3500
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: Macintosh
; OPERATING SYSTEM: System 7.0.1
; SOFTWARE: Microsoft Word 5.1a
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/659,473
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Paul D. Yasger
; REGISTRATION NUMBER: 37,477
; REFERENCE/DOCKET NUMBER:
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 708/937-2341
; TELEFAX: 708/938-2623
; TELEX:
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 150 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: genomic DNA (C. pneumoniae)
US-08-659-473-8

Query Match 10.6%; Score 146.8; DB 4; Length 150;
Best Local Similarity 98.7%; Pred. No. 9.8e-35;
Matches 148; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
Oy 366 tacctcaacatcactagctgacatcacaggtcttggagccctccagatgctgtcac 425
|||||
Db 1 TACCTCAACATCAGTACGTGACATGACAGGCTCTTTGGTGACCTCCAGGATGCTGTAC 60

Qy 426 taataaagatacagcgctactgtatgaggaacacgcaatcgtcgaggatgggaac 485
|||||
Db 61 TAATAAAGATACAGCGCTACTGTATGAGGAACCGCAATCGTGGGTGGGAAC 120
|||||
Qy 486 taagaatccgatgaataaagtggcgc 515
|||||
Db 121 TAAGAATCCGATGCGAGTTAAAGTTGGCGC 150
|||||

RESULT 2
US-08-232-463-14/c
; Sequence 14, Application US/08232463
; Patent No. 5670367
; GENERAL INFORMATION:
; APPLICANT: DORNER, F.
; APPLICANT: SCHEIFLINGER, F.
; APPLICANT: FALKNER, F. G.
; TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
; NUMBER OF SEQUENCES: 52
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 1800 Diagonal Road, Suite 500
; CITY: Alexandria
; STATE: VA
; COUNTRY: USA
; ZIP: 22313-0299
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/232.463
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/935.313
; FILING DATE:
; APPLICATION NUMBER: EP 91 114 300.6
; FILING DATE: 26-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 30472/114 IMMU
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703)836-9300
; TELEFAX: (703)683-4109
; TELEX: 899149
; INFORMATION FOR SEQ ID NO: 14:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 7218 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; IMMEDIATE SOURCE:
; CLONE: ptzgpt-fls
US-08-232-463-14

Query Match 3.9%; Score 54.4; DB 1; Length 7218;
Best Local Similarity 9.3%; Pred. No. 4.6e-06;
Matches 43; Conservative 218; Mismatches 199; Indels 0; Gaps 0;

Qy 631 aataaacaacgactgagcttctaagagatgcaagataaccacagtagtccacagga 690
|||||
Db 1481 AATTACTATCTATGCAAGTAGTTAAAGAGATAGAGAAATTTGGTACRRRRRRRR 1422
|||||
Qy 691 acgcctgaattgctcaatttttagttgatcacagatgtctacagcacacagatag 750
|||||
Db 1421 RRR 1362
|||||
Qy 751 aagagatggaatcgattggggatgatatttgcaggacagacgctagtgagctgta 810
|||||

Db 1361 RRR 1302
Qy 811 gaaatgctaaactaataacagataagcaacatagattcagctaaagcagcaatcgct 870
|||||
Db 1301 RRR 1242
|||||
Qy 871 acLgctaagacacaaatagctgaagctcagaaaaagtcccccagactctccaattctcaa 930
|||||
Db 1241 RRR 1182
Qy 931 gaagcggaacaaatggttaatacagcgctgagaagatcttaaaaaatacaaacctgcagat 990
|||||
Db 1181 RRR 1122
Qy 991 ggttctgatgttccaaatccagaaactacagttggaggtccccaagcaacaggaagtagt 1050
|||||
Db 1121 RRR 1062
Qy 1051 attggtagtagtctgtttccatgctgttagatgtagtctg 1090
|||||
Db 1061 AAGCTCCCTGCACCTGCAGCAAGCTCGGAATTAATTCG 1022

RESULT 3
US-08-072-070-1
; Sequence 1, Application US/08072070
; Patent No. 5476929
; GENERAL INFORMATION:
; APPLICANT: Briles, David E
; APPLICANT: Yother, Janet L
; APPLICANT: McDaniel, Larry S
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/072.070
; FILING DATE: 19930603
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/835.698
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/656.773
; FILING DATE: 15-FEB-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 521-0378
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rxl
; IMMEDIATE SOURCE:


```

; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/214,164
; FILING DATE: 17-MAR-1994
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/656,773
; FILING DATE: 15-FEB-1991
; CLASSIFICATION: 424
; ATTORNEY/AGENT INFORMATION:
; NAME: Berkstresser, Jerry W.
; REGISTRATION NUMBER: 22,651
; REFERENCE/DOCKET NUMBER: 6102-137
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 521-0813
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..1983
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 127..1983
; US-08-214-164-1

```

```

Query Match          3.6%; Score 50.6; DB 1; Length 2085;
Best Local Similarity 49.8%; Pred. No. 3.1e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

QY 745 atagagaagatggaatcgattggggtgcatttttcaggagacagaaacgctagtggga 804
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 ACAGACAAAGCGCGAAAGACGACGACGATAGATGATAGATGAAGCTAAGAAAGCGCAA 531

QY 805 gctgtagaataatgcttaataacaggtataagcaacatagattcagctaaagcagca 864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 GAAGAGGCAAAACCTAAATTTAATCTGTTTCGAGCAATGGTAGTTCCTGAGCCAGAGCAG 591

QY 865 atcgctactgtaagacacaaatagctgaagctcagaaaagttcccca---ctctcca 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 TTGGCTGAGACTAGAAAATAATCAGAAGAGCTAAGAAAGCTAAGAAAGCTTACTTAA 651

QY 922 attctcagaagcgggaacaaatgtaatacaggtcgagaaagatctttaaataatcaaa 981
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 652 AAACCTAGAAGAGCTAAGCAAAATTAAGAGAGGCTGAGAAAAGCTACTGAGCCAAA 711

QY 982 cctgcagatggtctgattgttcccaatccaggaacatcagttggaggtccaaagcaaaa 1041
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 712 CAAAAGTGGATGCTGAAGAGTCTCTCTCAAGCTAAATCGCTGAATGGAAATCAA 771

QY 1042 ggaagtattgtt 1054
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 772 GTTCATAGACTAG 784

```

```

RESULT 6
US-08-469-434-1
; Sequence 1, Application US/08469434
; Patent No. 5753463
; GENERAL INFORMATION:
; APPLICANT: Briles, David E

```

```

; APPLICANT: Yother, Janet L
; APPLICANT: McDaniel, Larry S
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL
; TITLE OF INVENTION: PROTEIN
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Shoemaker and Mattare, Ltd
; STREET: Suite 1203, 2001 Jefferson Davis Highway
; CITY: Arlington
; STATE: Virginia
; COUNTRY: U.S.A.
; ZIP: 22202-0286
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/469,434
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/072,065
; FILING DATE: 03 JUNE 1993
; APPLICATION NUMBER: US/07/835,698
; FILING DATE: 12-FEB-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/07/656,773
; FILING DATE: 15-FEB-1991
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (703) 415-0810
; TELEFAX: (703) 521-0378
; TELEX: LUKPAT WASHINGTON
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; ORGANISM: Streptococcus pneumoniae
; STRAIN: Rx1
; IMMEDIATE SOURCE:
; CLONE: JY2008
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; FEATURE:
; NAME/KEY: CDS
; LOCATION: join(127..1983, 1987..1992, 1996..2007, 2011
; LOCATION: ..2025, 2029..2031, 2035..2085)
; US-08-469-434-1

```

```

Query Match          3.6%; Score 50.6; DB 1; Length 2085;
Best Local Similarity 49.8%; Pred. No. 3.1e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

QY 745 atagagaagatggaatcgattggggtgcatttttcaggagacagaaacgctagtggga 804
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 472 ACAGACAAAGCGCGAAAGACGACGACGATAGATGATAGATGAAGCTAAGAAAGCGCAA 531

QY 805 gctgtagaataatgcttaataacaggtataagcaacatagattcagctaaagcagca 864
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 532 GAAGAGGCAAAACCTAAATTTAATCTGTTTCGAGCAATGGTAGTTCCTGAGCCAGAGCAG 591

QY 865 atcgctactgtaagacacaaatagctgaagctcagaaaagttcccca---ctctcca 921
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 592 TTGGCTGAGACTAGAAAATAATCAGAAGAGCTAAGAAAGCTAAGAAAGCTTACTTAA 651

```


CITY: Arlington
 STATE: Virginia
 COUNTRY: U.S.A.
 ZIP: 22202-0286
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC Compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent In Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/246,636
 FILING DATE: 20-MAY-1994
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/656,773
 FILING DATE: 15-FEB-1991
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 07/835,698
 FILING DATE: 12-FEB-1992
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/048,896
 FILING DATE: 20-APR-1993
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (703) 415-0810
 TELEFAX: (703) 415-0813
 TELEX: LUKPAT WASHINGTON
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 2085 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 HYPOTHETICAL: NO
 ANTI-SENSE: NO
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus pneumoniae
 STRAIN: Rx1
 IMMEDIATE SOURCE:
 CLONE: JV4313
 FEATURE:
 NAME/KEY: intron
 LOCATION: 1..2085
 FEATURE:
 NAME/KEY: CDS
 LOCATION: join(127..1984)
 US-08-246-636-1

	Query Match	3.6%;	Score 50.6;	DB 2;	Length 2085;
	Best Local Similarity	49.8%;	Pred. No. 3.1e-05;		
	Matches 156;	Conservative 0;	Mismatches 154;	Indels 3;	Gaps
QY	745	atagagaagatggaatacgcgattggggatgcataattttgcaggacagaaacgtagtggg	804		
Db	472	ACAGACAAGCCGCAAAAGCAGCACAGATATAAGATCATAGATTGAAGCTAAGAAGCGCAA	531		
QY	805	gctgtagaaaaatgctaacttaaacagataaagcaacatatagatttcagtaaagcagca	864		
Db	532	GAAGAGCGAAAACCTAAATTTTAATCTGTTCAGCAATGGTATGTTCTTGAGCGACGACG	591		
QY	865	atcgctactgtcaagcacacaatagctgaagctcagaaaaagttcccgca---ctctcca	921		
Db	592	TTCGCTGAGACTAAGAAAAAATTCAGNAGAAGCTAAACAAAAAGCACCGAAGTACTTACTAAA	651		
QY	922	attcttcaaagagcggaacaaaatggtaatcacaggctgagaagaatctttaaaatatcaaa	981		
Db	652	AAACTAGAAGAAGCTTAAGCAAAATTTAGAAGAGGCTGAGAAAAAAGCTACTGAAGCCCAA	711		
QY	982	ccttgcatagttgctgatgttcccaaatccagaaactacagtgtggaggttcccaagcaaca	1041		
Db	712	CAAAAAGTGGATGCTCGAAGAAGTCGCTCCTCAAGCTAAATCGCTGAATTTGGAAATCAAA	771		

```

Qy 1042 ggaagttagtattg 1054
Db      | | | | |
       772 GTTCATAGACTAG 784

RESULT 11
US-08/-247-491A-1
; Sequence 1, Application US/08247491A
; Patent No. 5965400
; GENERAL INFORMATION:
; APPLICANT: BRILES, David E.
; APPLICANT: YOTHER, Janet L.
; TITLE OF INVENTION: STRUCTURAL GENE OF PNEUMOCOCCAL PROTEIN
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FROMMER LAWRENCE & HAUG LLP
; STREET: 745 Fifth Avenue
; CITY: New York
; STATE: NY
; COUNTRY: USA
; ZIP: 10151
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/247,491A
; FILING DATE: 23-JUN-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOWALSKI, Thomas J.
; REGISTRATION NUMBER: 32,147
; REFERENCE/DOCKET NUMBER: 454312-2041
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-588-0800
; TELEFAX: 212-588-0500
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2085 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; ORIGINAL SOURCE:
; STRAIN: Streptococcus pneumoniae Rx1
; IMMEDIATE SOURCE:
; CLONE: JY2008
; FEATURE:
; NAME/KEY: intron
; LOCATION: 1..2085
; FEATURE:
; NAME/KEY: CDS
; LOCATION: Join(127..1983, 1987..1992, 1996..2007, 2011
; LOCATION: ..2025, 2029..2031, 2035..2085)
US-08/-247-491A-1

Query Match          3.6%; Score 50.6; DB 2; Length 2085
Best Local Similarity 49.8%; Pred. No. 3.1e-05;
Matches 156; Conservative 0; Mismatches 154; Indels

Qy 745 atagagaagaatggaaatgcgattggggatgcataatttttcaggacagacaacgcgtta
   ||| |||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 472 ACAGACAAGAACCACCGCAAAAGCAGCGACGATAAGATGATAGTGAAGCTAAGAACAAC
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Qy 805 gcctgtagaagaatgcctaatacaacagataaagcaaacatatagctcagctaaag
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Db 532 GAAGAGCAAAAACTAATTTAATAC TGTTCGAGCAATGGTAGTGTTCTGTAGCCACG
   ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | ||| | |||
Qy 865 atcgctactgctaagcacacaaaatagctggaagctcagaaaaagtttcccccca---ct

```



```

; EARLIER APPLICATION NUMBER: 08/312,949
; EARLIER FILING DATE: 1994-09-30
; EARLIER APPLICATION NUMBER: 08/246,636
; EARLIER FILING DATE: 1994-05-20
; EARLIER APPLICATION NUMBER: 08/048,896
; EARLIER FILING DATE: 1993-04-20
; EARLIER APPLICATION NUMBER: 07/835,698
; EARLIER FILING DATE: 1992-02-12
; EARLIER APPLICATION NUMBER: 07/656,773
; EARLIER FILING DATE: 1991-02-15
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH: 2086
; TYPE: DNA
; ORGANISM: Streptococcus pneumoniae
US-08-446-201-2

```

```

Query Match      3.6%; Score 50.6; DB 3; Length 2086;
Best Local Similarity 49.8%; Pred. No. 3.1e-05;
Matches 156; Conservative 0; Mismatches 154; Indels 3; Gaps 1;

QY 745 atagagaagatggaatgcattggggtgcatattttgcagacagacgcctagtgga 804
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 472 acagacaaagccgcaaaagcgcagcagataagatgatatgaagctaaagaaacgcgaa 531
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 805 gctgtagaaaaatgctaaatctaataaacagtagtaagcaacatagattcagctaaagcagca 864
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 532 gaagaggcaaaaactaaatttaactgttcgagcaatggtagttcctgagccagagcag 591
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 865 atcgctactgtcaagacacaaatagctgaagctcagaaaaagttcccca---ctctcca 921
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 592 ttggtgagactaaagaaaaatcagaagaagcctaaacaaagcaccagaaacttaactaaa 651
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 922 attctcaagaagcggaacaaatggttaatacaggtctgagaagatcttataaaatatcaaa 981
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 652 aaactagaagaagctaaagcaaaattagaagggctgagaaaaaagctactgaagccaaa 711
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 982 cctgcagatggttctgatttccaaatccaggaactacagttggaggctcccaagcaacaa 1041
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 712 caaaaagtggatgctgagaagtcgtctctcctcaagctaaatcgtgaattggaaaatcaa 771
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

QY 1042 ggaagtagtattg 1054
   | | | | |
Db 772 gtccatagactag 784

```

Search completed: February 7, 2002, 15:54:07
Job time: 361 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: February 7, 2002, 21:32:23 ; Search time 7419.31 seconds
(without alignments)
2011.762 Million cell updates/sec

Title: US-09-391-606-4
Perfect score: 1389
Sequence: 1 atgtgtaacccattgtgtcc.....aagttctatggatcgagct 1389

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 11351937 seqs, 5372889281 residues

Total number of hits satisfying chosen parameters: 22703874

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

- EST: *
- 1: em_estfun: *
 - 2: em_esthum: *
 - 3: em_estin: *
 - 4: em_estom: *
 - 5: em_estpl: *
 - 6: em_estba: *
 - 7: em_estro: *
 - 8: em_estov: *
 - 9: em_htc: *
 - 10: gb_estl: *
 - 11: gb_est2: *
 - 12: gb_htc: *
 - 13: gb_gss: *
 - 14: em_gss_fun: *
 - 15: em_gss_hum: *
 - 16: em_gss_inv: *
 - 17: em_gss_pln: *
 - 18: em_gss_pro: *
 - 19: em_gss_prod: *
 - 20: em_gss_vrt: *
 - 21: em_gss_other: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	45	3.2	861	13	AZ682538	ENT1ALTTF
C 2	45	3.2	906	13	AZ550193	ENTCZ54TR
C 3	44.4	3.2	484	10	AA314486	EST186534
C 4	42.8	3.1	856	13	BH132709	ENTNG59TF
C 5	42.6	3.1	460	10	AA672611	VC59B01.I
C 6	42	3.0	624	11	BF865301	963058F02
C 7	41.8	3.0	586	13	AZ632588	IM0487L17
C 8	41.8	3.0	850	10	AV406016	AV406016
C 9	41.6	3.0	762	10	BE661568	724 GmaxS
C 10	41.4	3.0	842	13	AZ536150	ENTDB92TF
C 11	41.4	3.0	845	13	AZ685553	ENTM43TR
C 12	41.4	3.0	876	13	AZ691916	ENTJK22TF

AZ525704	247Pbc05
AU075947	AU075947
BG291509	602385708
AZ676761	ENTHP32TF
BG893782	kt15c04.y
AQ227560	HS_2019_B
AL103838	Drosophil
AL067752	Drosophil
AL439765	T3 end of
AL065906	Drosophil
RI5357	Yf9c02.r1
BE783140	CIT-HSP-200
BG339626	602437485
BB454826	B8454826
BI394077	papin.pko
AZ667395	ENTLH15TR
AI218089	qh22b01.x
RO3975	pk13e11.r1
BE217675	MD0305 Me
AZ688493	ENTIO70TR
AZ548883	ENTES80TR
AZ693152	ENTLC06TF
AZ549589	ENTDD36TF
AZ525221	207Pbc07
BG602389	EST501479
BG124392	EST470038
AZ136778	SP_0175_A
AZ692954	ENTLG27TF
AL225985	Tetraodon
AL564621	AL564621
AL018519	F.rubripe
AV405991	AV405991

ALIGNMENTS

RESULT 1	AZ682538	861 bp	DNA	GSS	14-DEC-2000
LOCUS	ENT1ALTTF	Entamoeba histolytica	Sheared DNA	Entamoeba histolytica	
DEFINITION	genomic, DNA sequence.				
ACCESSION	AZ682538				
VERSION	AZ682538.1	GI:11819684			
KEYWORDS	GSS:				
SOURCE	Entamoeba histolytica.				
ORGANISM	Entamoeba histolytica				
REFERENCE	Eukaryota; Entamoebidae; Entamoeba.				
AUTHORS	Loftus, B., Van Aken, S. and Fraser, C.				
TITLE	Determination of clone end sequences from Entamoeba histolytica				
JOURNAL	HMI:IMSS sheared DNA library				
COMMENT	Unpublished (2000)				
	Contact: Brendan J Loftus				
	Department of Eukaryotic Genomics				
	The Institute for Genomic Research				
	9712 Medical Center Dr., Rockville, MD 20850, USA				
	Tel: 301 838 0200				
	Fax: 301 838 0208				
	Email: entaetigr.org				
	Clones are derived from the Entamoeba histolytica HMI:IMSS sheared				
	DNA library				
	Seq primer: M13-Forward				
	Class: shotgun				
	High quality sequence start: 36				
	High quality sequence stop: 813.				
	Location/Qualifiers				
FEATURES	1..861				
source	/organism="Entamoeba histolytica"				
	/strain="HMI:IMSS"				
	/db_xref="taxon:5759"				
	/clone_lib="Entamoeba histolytica Sheared DNA"				

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1978, Vol. 1, No. 2, pp. 1-10
3. The Importance of Parental Involvement	Journal of Educational Psychology	1985, Vol. 77, No. 3, pp. 1-10
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1990, Vol. 93, No. 1, pp. 1-10
5. The Role of the School in the Community	Journal of Educational Research	1995, Vol. 98, No. 1, pp. 1-10
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2000, Vol. 103, No. 1, pp. 1-10
7. The Role of the Teacher in the Classroom	Journal of Educational Research	2005, Vol. 108, No. 1, pp. 1-10
8. The Impact of Technology on Education	Journal of Educational Technology	2010, Vol. 3, No. 2, pp. 1-10
9. The Importance of Parental Involvement	Journal of Educational Psychology	2015, Vol. 87, No. 3, pp. 1-10
10. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	2020, Vol. 123, No. 1, pp. 1-10

FEATURES
SOURCE

Query Match 3.0%; Score 42; DB 11; Length 624;
Best Local Similarity 46.3%; Pred. NO. 2;

Qy 684 agggaaaacgcctgcaattgctcaatctttagttgatcagacagatgctacagcgacaca 743

QY 744 gatagagaagatggaatgcgattgggatgcataatttgcaggacagaaacgctaattgg 803

QY 804 agctgtagaaaaatgctaaatctataaacagtataagcaacatagattcagctaaagcagc 863

QY 864 aatcgc tactgcta agacacaa atagctga agctcaga aaaaagtt ccccgact ctcacat 923

924 tctccaagaagggaacaaatctgtaaacaggctgagaagatcttaaaaaatacaca 981
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000

500 501 502 503 504 505 506 507 508 509 510 511 512 513 514 515 516 517 518 519 520 521 522 523 524

AZ632588/c
LOCUS
AZ632588
586 bp
DNA
GSS
13-DEC-2000

AZ632588
 AZ632588.1
 GI:11754778

SOURCE	ORGANISM
house mouse.	Mus musculus

RESULT 10
AZ536150/c

RESULT	11	
AZ685553/c		
LOCUS		
DEFINITION		
ACCESSION		
KEYWORDS		
SOURCE		
ORGANISM		
REFERENCE		
AUTHORS		
TITLE		
JOURNAL		
COMMENT		
FEATURES		
source		
BASE COUNT		
ORIGIN		

```

Qy 1060 att 1062
Db 100 ATT 98

RESULT 12
LOCUS AZ691916
DEFINITION ENTJ22TF Entamoeba histolytica Sheared DNA Entamoeba histolytica genomic, DNA sequence.
VERSION AZ691916
KEYWORDS AZ691916.1 GI:11829182
SOURCE GSS.
ORGANISM Entamoeba histolytica.
REFERENCE Eukaryota; Entamoebidae; Entamoeba.
AUTHORS 1 (bases 1 to 876)
Loftus B., Van Aken, S. and Fraser, C.
TITLE Determination of clone end sequences from Entamoeba histolytica
HMI:IMSS sheared DNA library
UNPUBLISHED (2000)
CONTACT: Brendan J Loftus
DEPARTMENT OF Eukaryotic Genomics
THE INSTITUTE FOR GENOMIC RESEARCH
9712 MEDICAL CENTER DR., ROCKVILLE, MD 20850, USA
TEL: 301 838 0200
FAX: 301 838 0208
EMAIL: enta@tigr.org
CLONES ARE DERIVED FROM THE Entamoeba histolytica HMI:IMSS sheared DNA library
Seq primer: M13-Forward
Class: Shotgun
High quality sequence start: 16
High quality sequence stop: 801.
Location/Qualifiers
1. 876
/organism="Entamoeba histolytica"
/strain="HMI:IMSS"
/db_xref="taxon:5759"
/clone_lib="Entamoeba histolytica Sheared DNA"
/note="Vector: PHOS1; Site 1: Bst I; Constructed at The Institute for Genomic Research (TIGR), Rockville, MD.
Genomic DNA isolated from broth cultures of E. histolytica using a method described by Clark and Diamond (Clark, C.G., and Diamond, L.S. (1993) Entamoeba histolytica: a method for isolate identification. Exp. Parasitol. 77:450.). The DNA was mechanically sheared to give a tight size distribution (~2 kb). The v + i method used for the library construction is described in detail in Smith, H.O. and Venter, J.C. (Making small insert libraries for whole genome shotgun sequencing projects. In Genome Sequencing: A Practical Approach, eds. M. Vaudin and B. Barrell, Oxford University Press, 1999)."
134 g
80 c
348 a
BASE COUNT
ORIGIN

```

db 560 ATT 562

db 560 ATT 562

TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
 JOURNAL Unpublished (1999)
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: ATCC
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: LLAM10403 row: 0 column: 02
 High quality sequence start: 12
 High quality sequence stop: 594
 Location/Qualifiers
 1. .1000
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone="IMAGE:4514905"
 /clone_lib="NIH_MGC_93"
 /tissue_type="transitional cell papilloma, cell line"
 /lab_host="DH10B (phage-resistant)"
 /note="Organ: bladder; Vector: pCMV-SPORT6; Site_1: NotI;
 Site_2: SalI; Cloned unidirectionally; oligo-dr primed.
 Average insert size 1.7 kb. Library enriched for
 full-length clones and constructed by Life Technologies.
 Note: this is a NIH_MGC Library."

FEATURES

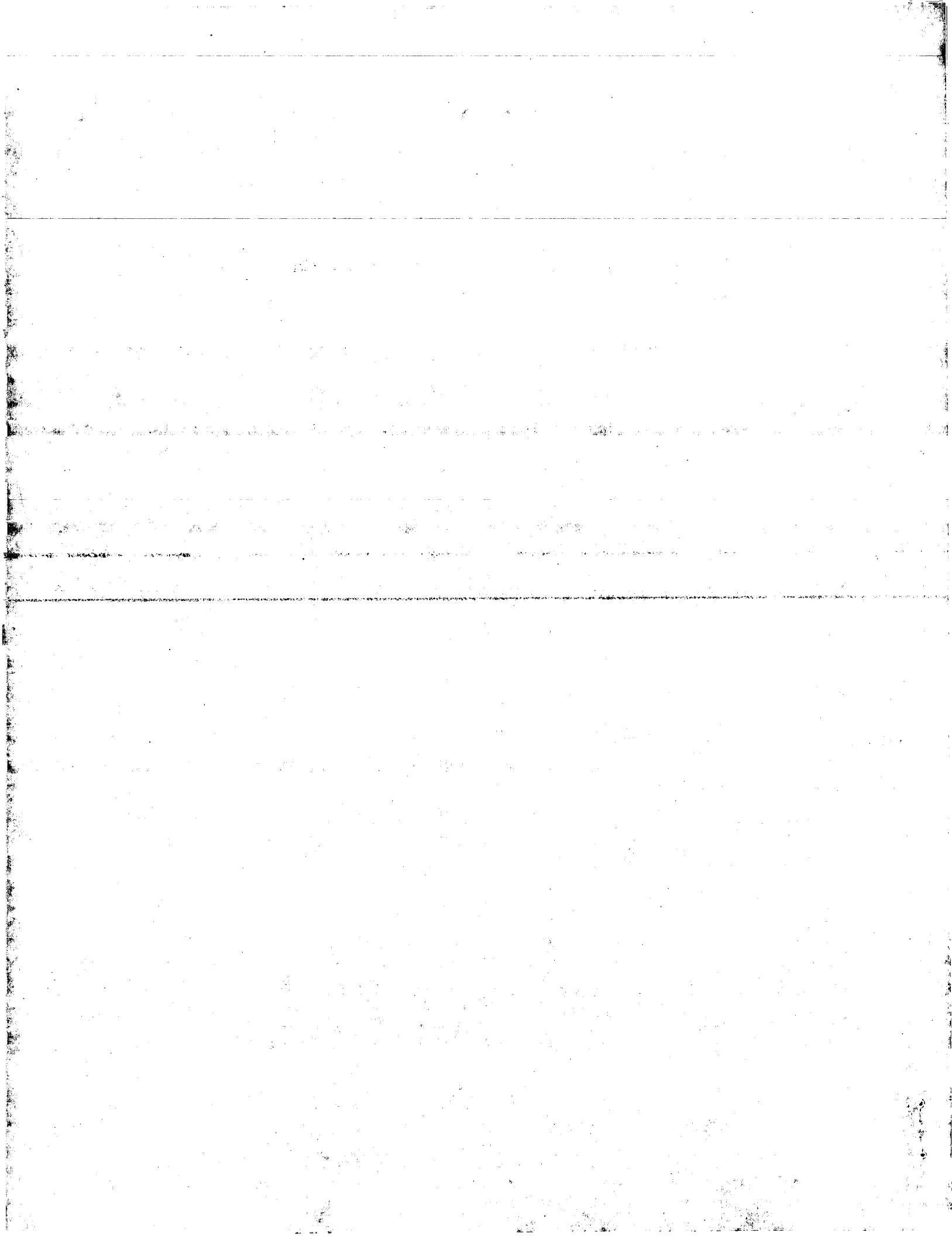
source

BASE COUNT 369 a 195 c 253 g 183 t
 ORIGIN

Query Match 3.0%; Score 41.2; DB 11; Length 1000;
 Best Local Similarity 49.1%; Pred. No. 3.8;
 Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;

Qy 750 gaaagatggaatcgattggggatgcataatttgcaggacagacgctagtgagctgt 809
 |||||
 Db 178 GAAAAAGAACGGGATGTTGATGGAGAAATGGTGCCTGTGGAGAAAAATGGTGAAGA 237
 |||||
 Qy 810 agaaaatgctaaatcctaagcagtaagcaacatagattcagctaaagcagcaatcgc 869
 |||||
 Db 238 AAACAAAGATGATGATATGTTTCAGCAGCAATTTGCCCTCAGAGCCTGTGGACATCTC 297
 |||||
 Qy 870 tactgctaagacacaaatagctgaagctcagaaaaaagtccccgactctccaattctca 929
 |||||
 Db 298 TACAGCAATGAGTGAACCGGCACTTGCTCAGAAAAAGAGTCAGTGAATGCATTGTATCT 357
 |||||
 Qy 930 agaagcggacaaatgtaatacacaggctgagaaagatcttaa 971
 |||||
 Db 358 TGAAGCCATGAGCATGTTTAATAGAGCTCAGGAAGGATTGA 399
 |||||

Search completed: February 7, 2002, 21:32:29
 Job time: 20663 sec



Db 443 THNPASEQTLQIAEALGL-----LVQEAFTDWTYGG-KKP-YDVGRFFKEDATHPE 491
 Qy 88 AKA-----TKTTL-NGKENLA---WFICGTTGLGRKAGDWSATVRYEVEALSVEI 135
 Db 492 ARKGEKWSDFDLRTMVERGKNNPAIFMWSIGNEIGEANGDAHSLATVK-RLVKV--IKDV 548
 Qy 136 DVSIGIRGNLLKFWFA-----QAIAANYDPKEANSFTNYKGFSAFY---MY 178
 Db 549 DKTRVTMGADKFRFGNGSGGHEKIADELDAVGNY-----SEDNKALRAKHPKWLII 602
 Qy 179 GITSLSFRAYGAYSKP-----ANDKLG-----SDFTFRKFDLGI 213
 Db 603 GSETSATRTRGSYYRPRELKHNSGPERNYEQSDYGNDRVGWGKTATASWTFDRDNAGY 662
 Qy 214 ISAF 217
 Db 663 AGQF 666

RESULT 2
 US-09-306-595C-7
 ; Sequence 7, Application US/09306595C
 ; Patent No. 6284506
 ; GENERAL INFORMATION:
 ; APPLICANT: HOSHINO, Tatsuo
 ; APPLICANT: OJIMA, Kazuyuki
 ; APPLICANT: SETOGUCHI, Yutaka
 ; TITLE OF INVENTION: ISOPRENOL PRODUCTION
 ; FILE REFERENCE: ISOPRENOL PRODUCTION
 ; CURRENT APPLICATION NUMBER: US/09/306,595C
 ; CURRENT FILING DATE: 1999-05-06
 ; PRIOR APPLICATION NUMBER: 98108210
 ; PRIOR FILING DATE: 1998-05-06
 ; NUMBER OF SEQ ID NOS: 43
 ; SOFTWARE: Patentin Ver. 2.1
 ; SEQ ID NO 7
 ; LENGTH: 1091
 ; TYPE: PRT
 ; ORGANISM: Phaffia rhodozyma
 US-09-306-595C-7

Query Match 6.9%; Score 81; DB 4; Length 1091;
 Best Local Similarity 22.4%; Pred. No. 2.5;
 Matches 48; Conservative 30; Mismatches 50; Indels 86; Gaps 12;
 Qy 8 WV--VEGI-----LNRLPKQFFVKCSVVDWNTFFVSETST-----TEKA 44
 Db 756 WLDSVEGMEVMAASPNSYRFRLOSIRKGMAGRSYIRLATSTGDAMNMNAGKGTAKA 815
 Qy 45 ATNAMKY-----KYCWQWLGVKHSQVPWINGKKPL-----YLYGAFLMNLAK 89
 Db 816 LETLSEYFPMOILALSGNYCIDK---KPSAINWIEGRKSVVAESVIPAIVKSVL-- 869
 Qy 90 ATKTT-----LNGKENLAWFTG---GTLGLR-----KAGDWSATV 122
 Db 870 --KTTVADLVNLNIRKNL---IGSAMAGSIGGFNAHSDILTSIFLATQDPAQNVESSM 924
 Qy 123 RYEVYEAAL-----SVPEIDVSGIGRGNLL 146
 Db 925 CMTLMEAVNDGKDLITCSMPAIECGTVGGGTF 958

RESULT 3
 US-08-822-445-12
 ; Sequence 12, Application US/08822445
 ; Patent No. 5952223
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaplan, Jerry
 ; APPLICANT: Perou, Charles
 ; APPLICANT: Moore, Karen
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
 ; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME

NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036/2711
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Diskette
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: DOS
 SOFTWARE: FastSeq Version 2.0
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/822,445
 FILING DATE: 21-MAR-1997
 CLASSIFICATION: 435
 ATTORNEY/AGENT INFORMATION:
 NAME: Coruzzi, Laura A.
 REGISTRATION NUMBER: 30,742
 REFERENCE/DOCKET NUMBER: 7853-062-999
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-790-9090
 TELEFAX: 212-869-8864
 TELEX: 66141 PENNIE
 INFORMATION FOR SEQ ID NO: 12:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 3672 amino acids
 TYPE: amino acid
 TOPOLOGY: unknown
 MOLECULE TYPE: protein
 US-08-822-445-12

Query Match 6.6%; Score 77.5; DB 2; Length 3672;
 Best Local Similarity 22.2%; Pred. No. 37;
 Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;
 Qy 31 NTFVPSSTTTEKAATNAMKYCWQWLGVKHSQVP-----WQHLVLTLYLOQPOGKRRIHCKISIWVSGQRKPDVT 1626
 Db 1585 NIFLPK-----WQHLVLTLYLOQPOGKRRIHCKISIWVSGQRKPDVT 1626
 Qy 79 YGAFLMNPLAKATKTTLNGKENLAW-FIGGTLGG-----LRKAGDWSATVRYEVEALSVP 133
 Db 1627 LDFML-----PRKTSLSDSNKTFCMIGHCLSSQEEFLQLAGKWL-----1667
 Qy 134 EIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAFYITDLSFRAYGAYS 193
 Db 1668 -----GNLLUFNGAKVGSQ-----EAFYLYACGNPHTSVMPCKYIG 1702
 Qy 194 KPAND 198
 Db 1703 KPVND 1707

RESULT 4
 US-822-445-10
 ; Sequence 10, Application US/08822445
 ; Patent No. 5952223
 ; GENERAL INFORMATION:
 ; APPLICANT: Kaplan, Jerry
 ; APPLICANT: Perou, Charles
 ; APPLICANT: Moore, Karen
 ; TITLE OF INVENTION: COMPOSITIONS FOR THE DIAGNOSIS
 ; AND TREATMENT OF CHEDIAK-HIGASHI SYNDROME
 ; NUMBER OF SEQUENCES: 32
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Pennie & Edmonds
 STREET: 1155 Avenue of the Americas
 CITY: New York
 STATE: New York
 COUNTRY: USA
 ZIP: 10036/2711

REGISTRATION NUMBER: 30,492
REFERENCE/DOCKET NUMBER: 34506.036
TELECOMMUNICATION INFORMATION:
TELEPHONE: 608-831-2100
TELEFAX: 608-831-2106
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-704B-17

Query Match 6.6%; Score 76.5; DB 2; Length 367;
Best Local Similarity 19.7%; Pred. No. 1.7; Indels 47; Gaps 7;
Matches 37; Conservative 24; Mismatches 80; Indels 47; Gaps 7;
QY 5 HYAWVVEGILNRLPKQFFVKSVVDWNTFVPSETSTTEKAATNAMKYKVCVWQWLVGRHS 64
Db 156 HHSFTRPTFDYQPK-----DWSWYNDNRVTA-----NYGWDWCWNNGNRR 198
QY 65 Q-----VPWINGCKKPLYLYGAFMLNPLAKATKTLINGKENLAWFISGGLGG---LRKAGD 117
Db 199 YMRMGVYVWGPKEFYIDGQLVRVYHYNATATKYNGTWYQYF--NAMNGQFPANNANG 256
QY 118 WSATVRYEVEALSYPEID-----VSGIGRGN-----LLKFWFAQAATAA 156
Db 257 YTAVTYTTSTTSYSPPTQAASNSNGISVIDPGNFGGAGFTKAMDIIINVEOXQWLAL 316
QY 157 NYDPKEAN 164
Db 317 NHTPSDAD 324

RESULT 7
US-09-107-755-17
Sequence 17, Application US/09107755
Patent No. 6001636
GENERAL INFORMATION:
APPLICANT: Knuth, Mark W.
APPLICANT: Knoche, Kimberly K.
APPLICANT: Selman, Susanne
APPLICANT: Hartnett, James R.
TITLE OF INVENTION: ISOLATED AGARASE ENZYMES FROM
TITLE OF INVENTION: FLAVOBACTERIUM SP. STRAIN NR19, CLONED GENES THEREFOR,
TITLE OF INVENTION: AND EXPRESSION THEREOF IN TRANSFORMED HOST CELLS
NUMBER OF SEQUENCES: 17
CORRESPONDENCE ADDRESS:
ADDRESSEE: Dewitt Ross & Stevens S.C.
STREET: 8000 Excelsior Drive, Suite 401
CITY: Madison
STATE: Wisconsin
COUNTRY: U.S.A.
ZIP: 53717-1914
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/107,755
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/655,704
FILING DATE: 03-JUN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Sara, Charles S.
REGISTRATION NUMBER: 30,492
REFERENCE/DOCKET NUMBER: 34506.036
TELECOMMUNICATION INFORMATION:

TELEPHONE: 608-831-2100
TELEFAX: 608-831-2106
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 367 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-107-755-17

Query Match 6.6%; Score 76.5; DB 3; Length 367;
Best Local Similarity 19.7%; Pred. No. 1.7; Indels 47; Gaps 7;
Matches 37; Conservative 24; Mismatches 80; Indels 47; Gaps 7;
QY 5 HYAWVVEGILNRLPKQFFVKSVVDWNTFVPSETSTTEKAATNAMKYKVCVWQWLVGRHS 64
Db 156 HHSFTRPTFDYQPK-----DWSWYNDNRVTA-----NYGWDWCWNNGNRR 198
QY 65 Q-----VPWINGCKKPLYLYGAFMLNPLAKATKTLINGKENLAWFISGGLGG---LRKAGD 117
Db 199 YMRMGVYVWGPKEFYIDGQLVRVYHYNATATKYNGTWYQYF--NAMNGQFPANNANG 256
QY 118 WSATVRYEVEALSYPEID-----VSGIGRGN-----LLKFWFAQAATAA 156
Db 257 YTAVTYTTSTTSYSPPTQAASNSNGISVIDPGNFGGAGFTKAMDIIINVEOXQWLAL 316
QY 157 NYDPKEAN 164
Db 317 NHTPSDAD 324

RESULT 8
US-08-961-083-46
Sequence 46, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8504
TELEFAX: (301) 309-8512
INFORMATION FOR SEQ ID NO: 46:
SEQUENCE CHARACTERISTICS:
LENGTH: 241 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear

; NAME: Oblon, No. 5871910man F.
 ;
 ; REGISTRATION NUMBER: 24,618
 ;
 ; REFERENCE/DOCKET NUMBER: 660-060-0 PCT
 ;
 ; TELECOMMUNICATION INFORMATION:
 ;
 ; TELEPHONE: (703) 413-3000
 ;

NAME: Oblon, No. 6013508man
REGISTRATION NUMBER: 24,618

NAME: Ob
REGISTRAT
REGISTRAT

TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
TELEX: 248855 OPAT UR
INFORMATION FOR SEQ ID NO: 19:
SEQUENCE CHARACTERISTICS:
LENGTH: 988 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-980-357-19

Query Match 6.3%; Score 73.5; DB 3; Length 988;
Best Local Similarity 23.0%; Pred. No. 16; Indels 5; Gaps 2;
Matches 17; Conservative 18; Mismatches 34; Indels 5; Gaps 2;

Qy 123 RYVEALSVPEIDVSGIGRNLKFWFAQIAAAYDPKANSFTNYKGFSA--YMYGI 180
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 234 RLEYRGMDLETVOISHLHRRLLQL---SRLGSRYPYAFRDFQENKRYSLTYLLQL 290
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Qy 181 TDSLFRAYGAYSK 194
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 291 TQELTDKAFEIHDR 304

RESULT 11
US-08-559-492-12
Sequence 12, Application US/08559492
Patent No. 5843884
GENERAL INFORMATION:
APPLICANT: Sims, Peter J.
TITLE OF INVENTION: C9 Complement Inhibitor
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: Patrea L. Pabst
STREET: 2800 One Atlantic Center, 1201 West Peachtree
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30309-3450

COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/559,492
FILING DATE: 15-NOV-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Pabst, Patrea L.
REGISTRATION NUMBER: 31,284
REFERENCE/DOCKET NUMBER: OMRf154
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404-873-8794
TELEFAX: 404-873-8795
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 561 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-559-492-12

Query Match 6.2%; Score 72.5; DB 2; Length 561;
Best Local Similarity 26.2%; Pred. No. 8.9; Indels 23; Gaps 6;
Matches 34; Conservative 20; Mismatches 53; Indels 23; Gaps 6;

Qy 84 MNPLAKA-TKTTINGKENLAWFIGTGLGRKAGDSATV-----RYEYVEA--- 129

Db 170 MDPLATPFEDNEYHYGLCDRWV-DGNTLTHYRK--PWNVAVLAYETKIDKNFRTEYEQM 236
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Qy 130 ---LSVPEIDVSGIGRNLKFWFAQIAAAYD---PKEANSFTNYKGFSAIYWGTD 183
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 227 QAFKSIIEETSFNANLALKFTPTTEAKASAEASPKNKSLDNDKGFSSKFQFSYKN 286
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Qy 184 LSFRAYGAYS 193
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 287 ETYQLFSLYS 296

RESULT 12

US-08-272-255-13
Sequence 13, Application US/08272255
Patent No. 5824859
GENERAL INFORMATION:
APPLICANT: Cashmore, Anthony R.
APPLICANT: Ahmad, Margaret
APPLICANT: Lin, Chentao
TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
TITLE OF INVENTION: Using the Same
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & No. 5824859ris
STREET: One Liberty Place, 46th floor
CITY: Philadelphia
STATE: PA
COUNTRY: USA
ZIP: 19103
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/272,255
FILING DATE: 08-JUL-1994
CLASSIFICATION: 800
ATTORNEY/AGENT INFORMATION:
NAME: Leary Ph.D., Kathryn
REGISTRATION NUMBER: 36,317
REFERENCE/DOCKET NUMBER: UPN-1795
TELECOMMUNICATION INFORMATION:
TELEPHONE: (215) 568-3100
TELEFAX: (215) 568-3439
INFORMATION FOR SEQ ID NO: 13:
SEQUENCE CHARACTERISTICS:
LENGTH: 475 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-272-255-13

Query Match 6.1%; Score 71.5; DB 2; Length 475;
Best Local Similarity 24.3%; Pred. No. 9.1; Indels 33; Gaps 7;
Matches 36; Conservative 19; Mismatches 60; Indels 33; Gaps 7;

Qy 55 VVQMLVGHKSQVPWINGOKKPLYLGAPLMN-----PLAKATKTLNGKENL---AW 103
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 127 LWDQLL--HSPDQILSGSGNYSYVGPFWKNWQAQPKPTVPATPTLVDSPEQLTAP 184
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Qy 104 FIGGTGLGRKAG-DWSATVRYEYVEALSVPEIDVSGIGRNLKFWFAQIAAAYDPKE 162
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 185 LLLSELPTLKQLGFDWDGCGFPVEGETAARIQEQ-----FCDRADIADYDQ- 231
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Qy 163 ANSFNTYKGFSAIYWGTDLSFRAYG 190
| | | : : : | | | : : | | | : : | | | : : | | | : : | | | : :
Db 232 -RNPPAEAGTS-----GLSPALKFGAIG 253

RESULT 13

PCT-US95-08565-13
; Sequence 13, Application PC/TUS9508565
; GENERAL INFORMATION:
; APPLICANT: Cashmore, Anthony R.
; APPLICANT: Ahmad, Margaret
; APPLICANT: Lin, Chentao
; TITLE OF INVENTION: Blue Light Photoreceptors and Methods of
; TITLE OF INVENTION: Using the Same
; NUMBER OF SEQUENCES: 22
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Woodcock, Washburn, Kurtz, Mackiewicz & Norris
; STREET: One Liberty Place, 46th floor
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/08565
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/272,255
; FILING DATE: 08-JUL-1994.
; ATTORNEY/AGENT INFORMATION:
; NAME: Leary Ph.D., Kathryn
; REGISTRATION NUMBER: 36,317
; REFERENCE/DOCKET NUMBER: UPN-1795
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (215) 568-3100
; TELEFAX: (215) 568-3439
; INFORMATION FOR SEQ ID NO: 13:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 475 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; PCT-US95-08565-13

Query Match 6.1%; Score 71.5; DB 5; Length 475;
Best Local Similarity 24.3%; Pred. No. 9.1;
Matches 36; Conservative 19; Mismatches 60; Indels 33; Gaps 7;
QY 55 VVOWLVGRKHSQVPWINGOKKPLYLGAFLLN-----PLAKATKTTTLNGKENL---AW 103
Db 127 LWDQLL--HSPDQLSSGNGPNYSVYGPFWKNNQAKPTTPVATPTELVDLSPEQLTAIAP 184
QY 104 FTGGTILGLRKAG-DWSATVREYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKE 162
Db 185 LLLSELPTLKQLGDFMDGFPVEPGETAIAIRLQE-----FCDAIADYDPO- 231
QY 163 ANSTNYKGFSAIYMYGTTDSLSEFRAYG 190
Db 232 -RNFFAEAGTS-----GLSPALKFGAIG 253

RESULT 14

US-09-540-245A-15
; Sequence 15, Application US/09540245A
; Patent No. 6270984
; GENERAL INFORMATION:
; APPLICANT: Goodman, Corey
; APPLICANT: Kid, Thomas
; APPLICANT: Brose, Katja
; APPLICANT: Tessier-Lavigne, Marc
; TITLE OF INVENTION: Modulating Robo: Ligand Interactions

FILE REFERENCE: B98-031-3
; CURRENT APPLICATION NUMBER: US/09/540,245A
; CURRENT FILING DATE: 2000-03-31
; PRIOR APPLICATION NUMBER: 60/065,544
; PRIOR FILING DATE: 1997-11-14
; PRIOR APPLICATION NUMBER: 60/081,057
; PRIOR FILING DATE: 1998-04-07
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 15
; LENGTH: 1395
; TYPE: PRT
; ORGANISM: Drosophila melanogaster
; US-09-540-245A-15
Query Match 6.1%; Score 71.5; DB 4; Length 1395;
Best Local Similarity 24.1%; Pred. No. 44;
Matches 26; Conservative 14; Mismatches 43; Indels 25; Gaps 5;
QY 68 WINGOKKPLYLGAFLLNPLAKATKTTTLNGKENLAWFIGG-----TLGGLRKAGDWS 119
Db 569 WAKSQEKP---GA--VGPIIGYTVVEYFSPDLOTGWIAAHRVGDQVTVISGLTPGTSYV 622
QY 120 ATVREYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFT 167
Db 623 FLVRAENTQGISVPS-----GLSNVIK-----TIEADFDAASANDLS 659
RESULT 15
US-08-816-346-56
; Sequence 56, Application US/08816346
; Patent No. 6127525
; GENERAL INFORMATION:
; APPLICANT: Crystal, Ronald G.
; APPLICANT: Falck-Pedersen, Erik
; APPLICANT: Gall, Jason
; APPLICANT: Koveshdi, Imre
; APPLICANT: Wickham, Thomas J.
; TITLE OF INVENTION: CHIMERIC ADENOVIRAL COAT PROTEIN AND
; TITLE OF INVENTION: METHODS OF USING SAME
; NUMBER OF SEQUENCES: 60
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: LEYDIG, VOIT & MAYER, LTD.
; STREET: TWO PRUDENTIAL PLAZA - 4900
; CITY: CHICAGO
; STATE: ILLINOIS
; COUNTRY: USA
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/816,346
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; REFERENCE/DOCKET NUMBER: 67167
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 312/616-5600
; TELEFAX: 312/616-5700
; TELEX: 25-3533
; INFORMATION FOR SEQ ID NO: 56:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 967 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-816-346-56

Query Match 6.0%; Score 70; DB 3; Length 967;
Best Local Similarity 20.7%; Pred. No. 38;
Matches 37; Conservative 25; Mismatches 61; Indels 56; Gaps 8;
Qy 56 WQWLVGKHSQVPWINGOKKPLYLY-----GAFLMNPLAKATKTTLNGKENLAWFIGG 107
Db 693 WAFTRLTKETPSLGSQVDPYTYSGSIPYLDGTFYLNHTFKKVAIT----- 739
Qy 108 TLGGLRKAGDWSATVRYEVEALSP---EIDVSGIGRG-----NLLKFWFAQAIANY 158
Db 740 -----EDSSVSWFGNDRLLTPNEFEIKRSVDGEGYNVAQCNTKDWFLVQMLANY 789
Qy 159 DPKEANSFTNYKGFSAIYMYGITSLSFRAYGAYS--KPANDKLGSDFTPRKF-DLGII 214
Db 790 N-----IGYQGFY-----IPESYKDRWYSFFRNFQPMRSRQVVDTKYKEYQQVGIL 835

Search completed: February 7, 2002, 21:36:13
Job time: 20482 sec

Db 386 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 2

F86581
CHLPN 76 kDa homolog_2 (CT623) [imported] - Chlamydomophila pneumoniae (strain J138)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 23-Mar-2001
C:Accession: F86581
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.; Ise
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of Chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: F86581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-422 <STO>
A:Cross-references: GB:BA000008; NID:g8979101; PIDN:BAA98936.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
C:Superfamily: Chlamydia trachomatis hypothetical protein CT623

Query Match 99.7%; Score 1162; DB 2; Length 422;
Best Local Similarity 99.5%; Pred. No. 3.6e-97;
Matches 216; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 60
Db 206 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 265
Qy 61 GKHSQVPWINGOKKPLYLYGAFLNPLAKATKTLNGKENLAWFTGGTGLGRKAGDWSA 120
Db 266 GKHSQVPWINGOKKPLYLYGAFLNPLAKATKTLNGKENLAWFTGGTGLGRKAGDWSA 325
Qy 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYNGI 180
Db 326 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYNGI 385
Qy 181 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217
Db 386 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 422

RESULT 3

I40729
hypothetical 76k protein - Chlamydomophila pneumoniae (strain AR39)
C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae
C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000
C:Accession: I40729
R:Perez-Melgosa, M.; Kuo, C.
Infect. Immun. 62, 880-886, 1994
A:Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76-kilobase protein.
A:Reference number: I40729; MUID:94156481
A:Accession: I40729
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-715 <RES>
A:Cross-references: GB:L23921; NID:g435961; PIDN:AAA23117.1; PID:g435962
A:Experimental source: strain AR-39
C:Comment: This is the hypothetical translation of a sequence that was reported as two s

Query Match 99.1%; Score 1156; DB 2; Length 715;
Best Local Similarity 99.5%; Pred. No. 2.4e-96;
Matches 215; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 60
Db 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFVPSETSTTEKAATNAMYKVCWQWLV 60
Qy 61 GKHSQVPWINGOKKPLYLYGAFLNPLAKATKTLNGKENLAWFTGGTGLGRKAGDWSA 120

Db 61 GKHSQVPWINGOKKPLYLYGAFLNPLAKATKTLNGKENLAWFTGGTGLGRKAGDWSA 120
Qy 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYNGI 180
Db 121 TVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAALYNGI 180
Qy 181 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216
Db 181 TDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 216

RESULT 4

H71490
hypothetical protein CT623 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
N:Alternate names: chlpn 76kDa homolog CT623
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 28-Jul-2000
C:Accession: H71490
R:Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitch
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia
A:Reference number: A71570; MUID:99000809
A:Accession: H71490
A:Molecule type: DNA
A:Residues: 1-446 <ARN>
A:Cross-references: GB:AE001333; GB:AE001273; NID:g3329068; PIDN:AAC68227.1; PID:g33
A:Experimental source: serotype D, strain UW-3/Cx
C:Comment: This sequence was originally identified as homologous to part of a sequen
PIR:E72042).
C:Genetics:
A:Gene: CT623
C:Superfamily: Chlamydia trachomatis hypothetical protein CT623

Query Match 66.7%; Score 778; DB 2; Length 446;
Best Local Similarity 63.8%; Pred. No. 1.6e-62;
Matches 148; Conservative 26; Mismatches 42; Indels 16; Gaps 4;

Qy 1 MTKKHAWVVEGILNRLPKQFFVKCSVVDWNTFV-----PSETSTTEKAATNAMYKVCV 55
Db 216 MAEREYAWVVEAILNKLPGNFVVKTSVVDWNTLAKTNDPADASAAQPAKPT-KYDYLV 274
Qy 56 WQWLVGKHSQVPWINGOKKPLYLYGAFLNPLAK-----ATKTT--LNGKENLAWFI 105
Db 275 WQWLVGKSTAMPFNGQTKNLYTYGAVLFNPLAETPENWKQSTTPTTKITNGKENHAWFI 334
Qy 106 GTTGLGRKAGDWSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANS 165
Db 335 GCSLGGVRRAGDWSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPKEANS 394
Qy 166 FTNYKGFSAALYNGITDSLFRAYGAYSKPANDKLGSDFTFRKFDLGIISAF 217
Db 395 FTNYKGVSYQFVGMGLTDSVFRAYAAKSPANDNLGSDFTYRKYDGLGISSE 446

RESULT 5

G69723
transketolase (EC 2.2.1.1) tkt - Bacillus subtilis
C:Species: Bacillus subtilis
C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 20-Jun-2000
C:Accession: G69723; S57401
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Be
C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;
A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrati,
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Ga
iech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo,
Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardin
A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Ma
Y. M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portet
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scan
A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; S

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K. A:Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033
A:Accession: G69723
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-667 <KUN>
A:Cross-references: GB:299113; GB:AL009126; NID:g2634090; PIDN:CAB13673.1; PID:g2634173
A:Experimental source: strain 168
R:Schlott, T.; von Wachenfeldt, C.; Hederstedt, L.
submitted to the EMBL Data Library, June 1995
A:Reference number: S57401
A:Accession: S57401
A:Molecule type: DNA
A:Residues: 561-667 <SCH>
A:Cross-references: EMBL:X87845; NID:g870921; PIDN:CAA61113.1; PID:g870922
C:Genetics: tkt
C:Superfamily: transketolase; thiamin pyrophosphate-binding domain homology
C:Keywords: transferase
F:145-195/Domain: thiamin pyrophosphate-binding domain homology <TPB>

Query Match 7.9%; Score 92; DB 2; Length 667;
Best Local Similarity 22.9%; Pred. No. 1.9; Mismatches 68; Indels 68; Gaps 12;
Matches 50; Conservative 32; Mismatches 68; Indels 68; Gaps 12;
QY 1 MYKKHYAVVVEGILNRLPKQFFVKCSVDWNTFVPSSTTEKAATNAMKYY-----C 54
DB 273 LTKAYATYE-----EDFVPSVEYEHFAVAVKGEKKEQWNAQFAKYKEVPEL 325
QY 55 VQWLVGHSQVPWINGOKKPLYLYGAFILNPLAKATKTLNG-KENLAWFTGGT--LGG 111
DB 326 AQOLELATKGLPKDWDQVPEYKGSLS--RASSGEVLNGLAKKIPFFVGGADLAG 383
QY 112 -----LRKAGDSATVRYEVEALSVPEDVSGIGRGNLLKFW-----FAQIAAANYDPK 161
DB 384 SNKTTIKNAGDTA-----VDYSG-----KNFVGFVREFAMGAALN---- 419
QY 162 EANSFNTYKGS---ALYMGITDLSFRAYGAYSKPA 196
DB 420 -----GMALHGLRVFGGT-----FFVFSYLRPA 444

RESULT 6
T17672
chitinase-like protein - *Chlorella* virus PBCV-1
C:Species: *Chlorella* virus PBCV-1
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T17672
R:Graves, M.V.; Van Etten, J.L.
submitted to the EMBL Data Library, May 1999
A:Reference number: Z18806
A:Accession: T17672
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-830 <GRA>
A:Cross-references: EMBL:U42580; NID:g4028896; PIDN:AAC96549.2
A:Experimental source: specific host *Chlorella* strain NC64A
C:Genetics:
A:Note: A181/182R

Query Match 7.6%; Score 88.5; DB 2; Length 830;
Best Local Similarity 22.4%; Pred. No. 5; Mismatches 44; Indels 69; Gaps 6;
Matches 38; Conservative 19; Mismatches 44; Indels 69; Gaps 6;
QY 4 KHYAVVVEGILNR-----LPKQFFVKCSVDWNTFVPSSTT-----EKAATNAMKY 51
DB 50 ENFTWFSEGLVRKGNKVTMIPK-----DWNNSIPAGTKIIPGCVKALPNLKY 100

QY 52 KYCVQWLVGHSQVPWINGOKKPLYLYGAFILNPLAKATKTLNGKENLAWFIGGTLGG 111
DB 101 NQIL--PLVGK-----DP-----S 112
QY 112 LRKAGDSATVRYEVEALSVPEDVSGIGRGNLLKFWFAQIAAANYDPK 161
DB 113 LAKRGKWSKAVAPYVDACAFPTDPLPAISKASGLKFFTLAPITADSNK 162
RESULT 7
A44054
orf1 protein - *Junonia coenia* densovirus
C:Species: *Junonia coenia* densovirus
C:Date: 03-Feb-1994 #sequence_revision 03-Feb-1994 #text_change 08-Oct-1999
C:Accession: A44054
R:Dumas, B.; Jourdan, M.; Pascaud, A.M.; Bergoin, M.
Virology 191, 202-222, 1992
A:Title: Complete nucleotide sequence of the cloned infectious genome of *Junonia coenia* densovirus
A:Reference number: A44054; MUID:93033112
A:Accession: A44054
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-810 <DUM>
A:Cross-references: GB:S47266; NID:g257675; PIDN:AAB23698.1; PID:g257676

Query Match 7.2%; Score 84; DB 2; Length 810;
Best Local Similarity 25.8%; Pred. No. 12; Mismatches 61; Conservative 22; Mismatches 59; Indels 94; Gaps 18;
Matches 61; Conservative 22; Mismatches 59; Indels 94; Gaps 18;
QY 2 TKKHYAVVVEGILNRLPKQFFVKCSVDWNTFVPSSTTEKAATNAMKYYCVWQWLVG 61
DB 413 TKSH-KEMIFGLN-----NVIGPTGTGT---AVNRL-ITTCI----- 446
QY 62 KHSQVPWINGOKKPLYLYG-----FLNPLA-----KATK-TTLN 96
DB 447 --AEIPW--QKPLYMNQSEFDLPPGSRVVECNVKVIFRTNRIAFETSSATKQATIN 501
QY 97 GRENLAWFIGGTLGLRKAGDSATVRYEVEALSVPEDVSGIGRGNLLKFWFAQIA-- 154
DB 502 QISNLQ-----TAVGLNKLK-G-----GIDR-SPTAQSDQPMIP 534
QY 155 ----AANYDPKBEANSFNTYKGSALYMGITDLSFRAYGAYSKPANGLGSDFTF 206
DB 535 TATSAPKYEPP--ITGTCYGRGMIADY-YG-ADSTNDAAFGNAGNYPHQVGS-FTF 585

RESULT 8
F71062
hypothetical protein PH1196 - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 20-Jun-2000
C:Accession: F71062
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; S. M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Og. DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic archaeon, *Pyrococcus horikoshii* strain OT3
A:Reference number: A71000; MUID:98344137
A:Accession: F71062
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-743 <KAW>
A:Cross-references: GB:AP000005; NID:g3236132; PIDN:BAA30296.1; PID:g3257613
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBank
C:Genetics:
A:Gene: PH1196
C:Superfamily: *Pyrococcus horikoshii* hypothetical protein PH1196

Query Match 7.1%; Score 83; DB 2; Length 743;
Best Local Similarity 18.5%; Pred. No. 14; Mismatches 42; Conservative 35; Mismatches 76; Indels 74; Gaps 9;
Matches 42; Conservative 35; Mismatches 76; Indels 74; Gaps 9;

QY 9 VVEGILNRLPKQFFVKCSVDWNTFVSETSTTEKAATNAMKYKVCVWQVLVGHKSQVPW 68
 Db 314 IIEELFNAPKQFEIYAEVSGYREI-----YYY-----NNGQW 346
 QY 69 INQOKKPLYLYGAFLMNPLAKATKTTTLNGKENLAWFIGTGLGRLKAGDWSATVRYEVE 128
 Db 347 IFDKEPF-----SAKYTTLSGYRHIG-----GGYVVGFGAGSLPADYKD 387
 QY 129 A-----LSVPEDVDSIGGRGNLLKFW-----FAQAIANYPDKPEANSFTN 168
 Db 388 APEAAKFTAFQTKVSGASSNLIARLWSAPLHMLNATVAYREYVQANPGYDFPSPSIIDE 447
 QY 169 YKGFSAFYWTGTDLSFRAGYASKPAN--DKLGSDFTFKFDLGI 213
 Db 448 ----ARIVVYRTIDKGTWEL--AVAIPLSGVNTNLGDLTGTGRRLDASV 488

RESULT 9
 A54760
 translation elongation factor eEF-1 alpha chain - Trypanosoma brucei
 C:Species: Trypanosoma brucei
 C:Date: 28-Oct-1994 #sequence_revision 21-May-1999 #text_change 02-Feb-2001
 C:Accession: A54760; C49394
 R:Kaur, K.J.; Ruben, L.
 J. Biol. Chem. 269, 23045-23050, 1994
 A:Title: Protein translation elongation factor-lalpha from Trypanosoma brucei binds caln
 A:Reference number: A54760; MUID:94364999
 A:Accession: A54760
 A:Status: preliminary
 A:Molecule type: mRNA
 A:Residues: 1-239, 'TCN', 243-449 <KAU>
 A:Cross-references: GB:U10562; NID:9507782; PIDN:AAA57476.1; PID:9507783
 R:Baldau, S.L.; Palmer, J.D.
 Proc. Natl. Acad. Sci. U.S.A. 90, 11538-11562, 1993
 A:Title: Animals and fungi are each other's closest relatives: congruent evidence from m
 A:Reference number: A49394; MUID:94089672
 A:Accession: C49394
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 18-415 <BAL>
 A:Cross-references: GB:L25868; NID:9438909; PIDN:AAA16602.1; PID:9438910
 C:Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog
 C:Keywords: calmodulin binding; GTP binding; nucleotide binding; P-loop
 F:8-156/Domain: translation elongation factor Tu homology <ETU>
 F:14-21/Region: nucleotide-binding motif A (P-loop)
 F:153-156/Region: GTP-binding NXKD motif

Query Match 7.1%; Score 82.5; DB 2; Length 449;
 Best Local Similarity 21.6%; Pred. No. 8.3;
 Matches 42; Conservative 26; Mismatches 71; Indels 55; Gaps 9;
 QY 19 KQFFVKCSYVD-----NNTFVPSSTSTTEKAATNAMKYKVCV---WQW---LVGKH 63
 Db 146 KQWVCCNKMDDKTVNGYQERYDEIVKEVSAYIKKGVYVNEKVRFPVPSGQGDNMIEKS 205
 QY 64 SQVPWINGQKKPLYLYGAFLMNPLAKATKTTTLNGKENLAWFIGTGLG-----GLRK 114
 Db 206 EKMPWYKGG---PTLLEALMDPEPVRPSDKPLRLPLQDYKIGG-IGTVFVGRVETGVNMK 261
 QY 115 AGD-----WSATVRYEVE-----ALSVPEDVDSIGGRGNLLKFWFAQA 153
 Db 262 PGDVVTFAPANVTTEVKSTEMHHEQLAEATPGDNVGFNVKNSVKDIRGNVC-----G 315
 QY 154 IAANYDPKEANSFT 167
 Db 316 NTKNDPPKEAADFT 329

RESULT 10
 G83558
 hypothetical protein PA0696 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa
 C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
 C:Accession: G83558
 R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.
 adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.;
 Lory, S.; Olson, M.V.
 Nature 406, 959-964, 2000
 A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic p
 A:Reference number: A82950; MUID:20437337
 A:Accession: G83558
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-568 <STO>
 A:Cross-references: GB:AE004505; GB:AE004091; NID:9946572; PIDN:AAG04085.1; GSPDB:G
 A:Experimental source: strain PA01
 C:Genetics:
 A:Gene: PA0696

Query Match 7.1%; Score 82.5; DB 2; Length 568;
 Best Local Similarity 23.4%; Pred. No. 11;
 Matches 29; Conservative 19; Mismatches 39; Indels 37; Gaps 5;
 QY 102 AWFITGTLGG---LRKAGDWSATVRYEVEALSVPEDVDSIGGRGNLLKFWFAQAIANY 158
 Db 474 AWMLOFTLGNALMDRAGDQVFAAYKYIQPDALP----- 508
 QY 159 DPKEANSF---TNYKGFSAFYWTGTDLSFRAGYASKPANLKL-GSDTFKFDLGI 213
 Db 509 DGFNDSTFHLGGTNAKG-----YILGASYGDFKRVYGTARMLSSDEVYGAFFSDIVLQLEV 564
 QY 214 ISAF 217
 Db 565 NTRF 568
 RESULT 11
 T18042
 ice nucleation protein homolog A540L - Chlorella virus PBCV-1
 C:Species: Chlorella virus PBCV-1
 C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 17-Nov-2000
 C:Accession: T18042; E46352
 R:Graves, M.V.; Van Etten, J.L.
 submitted to the EMBL Data Library, May 1999
 A:Reference number: Z18806
 A:Accession: T18042
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-1176 <GRA>
 A:Cross-references: EMBL:U42580; NID:94028896; PIDN:AAC96907.1
 R:Schuster, A.M.; Graves, M.; Korth, K.; Ziegelbein, M.; Brumbaugh, J.; Grone, D.; M
 virology 176, 515-523, 1990
 A:Title: Transcription and sequence studies of a 4.3-kbp fragment from a ds-DNA euka
 A:Reference number: A46352; MUID:90266467
 A:Accession: E46352
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 885-1176 <SCH>
 A:Cross-references: GB:M33758; NID:9323370; PIDN:AAA66404.1; PID:9807598
 C:Genetics:
 A:Note: A540L

Query Match 7.0%; Score 82; DB 2; Length 1176;
 Best Local Similarity 23.6%; Pred. No. 29;
 Matches 43; Conservative 26; Mismatches 61; Indels 52; Gaps 10;
 QY 59 LVGRHSQVPWINGQKKPLYLYGAFLMNPLAKATKTTTLNGKENLAWFIG----- 106
 Db 179 IIGPHAGENLNGNE--LVLIIGAY-----AGQNLNGSRDV--FVGYKAGQNAQLESD 227
 QY 107 ----GTLGGLR-----KAGDWSATVRYEVEALSVPEDVDSIGGRGNLLKFWF 150

Db 228 SVAIGRCAGLNTIGSNNVWVGSGSFPATK-----ETVAIGA--KAGYNAGNSLQSVF 281
Qy 151 AQAIAANYDPKREANSFTNYKGFSAIYMGVITDLSFRAYGAYSPANDKLGSDTFERKFD 210
Db 282 V-GRSAGYNGGTSTFGYEA-----GYGATDSSLYNTFFVGY-TGTNRTGTSFNTMGAD 334
Qy 211 LG 212
Db 335 AG 336

RESULT 12

JC5117
translation elongation factor eEF-1 alpha - Trypanosoma cruzi
C:Species: Trypanosoma cruzi
C:Date: 02-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 02-Feb-2001
C:Accession: JC5117
R:Billaut-Mulot, O.; Fernandez-Gomez, R.; Loyens, M.; Ouaisi, A.
Gene 174, 19-26, 1996
A:Title: Trypanosoma cruzi elongation factor 1-alpha: Nuclear localization in parasites
A:Reference number: JC5117; MUID:97017123
A:Accession: JC5117
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-449 <BIL>
A:Cross-references: GB:I76077; NID:gi929444; PIDN:AAC01751.1; PID:gi929445
C:Comment: This protein plays roles in genetic control of cell death and nuclear process
C:Genetics:
A:Gene: EF-1alpha
C:Superfamily: translation elongation factor Tu; translation elongation factor Tu homolog
C:Keywords: GTP binding; nucleotide binding; P-loop
F:8-156/Domain: translation elongation factor Tu homology <ETU>
F:14-21/Region: nucleotide-binding motif A (P-loop)
F:162-181/Region: actin binding #status predicted
F:230-266/Domain: tRNA binding #status predicted <TRN>

Query Match 7.0%; Score 81.5; DB 2; Length 449;
Best Local Similarity 21.6%; Pred. No. 10;
Matches 42; Conservative 25; Mismatches 72; Indels 55; Gaps 9;
Qy 19 KOFFVKCSVDP-----WNFTVPSETSTTEKAATNAMYKVCV---NQW--LVCKH 63
Db 146 KOMVCCNMDDKSNFAQERYDEIVKEVSAYLKKRGVNVKRFPIPIGSGOGNMDIKS 205
Qy 64 SQVPMINGQKPLYLGYAFLNPLAKATKTTLNKKNLAWFIGTGLG-----GLRK 114
Db 206 ENMPYKG---PTLLEALDMLPEPPVRPSDKPLRLPLQDVYKIGG-IGTVPVGRVETGTMK 261
Qy 115 AGD-----WSATVRYEVE-----ALSVPEIDVSGIGRGNLLKFWFAQA 153
Db 262 PGDVVTFAPANVTTEKVSIMHHEQLAEATPGDNVGNVKNVSKVDIRRGVNC-----G 315
Qy 154 IAANYDPKREANSFT 167
Db 316 NSKNDPPKEAADFT 329

RESULT 13

S74663
peptide methionine sulfoxide reductase msrA - Synechocystis sp. (strain PCC 6803)
N:Alternate names: hypothetical protein slr1795
C:Species: Synechocystis sp.
A:Variety: PCC 6803
C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S74663
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimpou, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis
s.
A:Reference number: S74322; MUID:97061201

A:Accession: S74663
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-214 <KAN>
A:Cross-references: EMBL:D90900; GB:AB001339; NID:gi1651768; PIDN:BAAL6815.1; PID:gi16
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Genetics:
A:Gene: msrA
A:Start codon: GTG
C:Superfamily: peptide methionine sulfoxide reductase

Query Match 6.9%; Score 80.5; DB 2; Length 214;
Best Local Similarity 24.6%; Pred. No. 5.1;
Matches 29; Conservative 16; Mismatches 34; Indels 39; Gaps 6;
Qy 80 GAFLMNLAKA-----TKTTLNCKENLAWFIGTGLGLRK----- 114
Db 8 GSFLISPSKVIPOVDINPVSTTARTEK-AVFAGGCGFWGLEAMFEFVRGVKDVQGY 66
Qy 115 AGDWSATVRY-----EYVEALSVPEDIVSGIGRGNLLKFWFAQAIAANYDPKEAN 164
Db 67 SGGTEATANYARVSGGTDHAESIEI-VYDPAQVSYGELLKIFF-----SVGHDPQV 119
RESULT 14
D70354
threonine synthase (EC 4.2.99.2) thrC1 [similarity] - Aquifex aeolicus
C:Species: Aquifex aeolicus
C:Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 26-May-2000
C:Accession: D70354
R:Decker, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.
V.
Nature 392, 353-358, 1998
A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.
A:Reference number: A70300; MUID:98196666
A:Accession: D70354
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-409 <AQF>
A:Cross-references: GB:AE000698; NID:gi2983224; PIDN:AAC06822.1; PID:gi2983225; GB:AE0
A:Experimental source: strain VF5
C:Genetics:
A:Gene: thrC1
C:Superfamily: threonine dehydratase
C:Keywords: carbon-oxygen lyase; phosphoprotein; pyridoxal phosphate
F:109/Binding site: pyridoxal phosphate (Lys) (covalent) #status predicted

Query Match 6.9%; Score 80.5; DB 2; Length 409;
Best Local Similarity 23.2%; Pred. No. 11;
Matches 33; Conservative 20; Mismatches 54; Indels 35; Gaps 6;
Qy 78 LYGAFLMNLAKATKTTLNKKNLAWFIGTGLG-----GLRK-AGDWSATVRYEVEAL 130
Db 164 IIGSLVFNPTVVAVEGNYDDVNRCLSEANDLGWAFVNIINIRPYPAESKTLAEVAQL 223
Qy 131 S--VPEIDVSGIGRGNL-LKFWFAQAIAANYDPKEANSFTNYKGFSAIYMGVITDLSLFR 187
Db 224 GWRAPDVVVVAPAAAGSLYTKIW-----KGFNELKVLGLIDEVKTR 263
Qy 188 AYGAYSKRPND-----KLGSDP 204
Db 264 MYGAQAGCCSPIAQNAREGRDF 285

RESULT 15

T35670
hypothetical protein SC7B7.02 SC7B7.02 - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C:Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C:Accession: T35670
R:Murphy, L.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, November 1997

A:Reference number: Z21586

A:Accession: T35670

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-469 <MUR>

A:Cross-references: EMBL:AL009199; PIDN:CAAL5782.1; GSPDB:GN000070; SCOEDB:SC7B7.02

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCOEDB:SC7B7.02

```

Query Match          6.9%; Score 80.5; DB 2; Length 469;
Best Local Similarity 24.2%; Pred. No. 13;
Matches 45; Conservative 20; Mismatches 68; Indels 53; Gaps 6;

Qy 10 VEGILNRLPKOFFVKCSVVDNNTFVPSETSTTEKAAATNAMKYKVCWOWLVGKHQVPMI 69
Db 260 VGVK-----VNETVVKQMFVPGGGGTQFQKA-----QAIS 293

Qy 70 NGQKKPLYLYGAFLMNLAKATKT-----TLNGKENLAWFIGGTLGGLRKAGDWS 119
Db 294 NDQKALLYPSGGWIENEMKKATKADFQMTGIPSMITLTDKPALPY-----EALRAAG-- 345

Qy 120 ATVRYEYVEALSVPEIDVSGIGRGNLLKFWFAQAIAANYDPKEANSFTNYKGFSAIYMG 179
Db 346 -----EPFIVPKQGNPAGGKREVLRLAMLSKAAANFS-KTKLAPTIVKGTVPADGCG 396

Qy 180 ITDLSL 185
Db 397 STALVS 402

```

Search completed: February 7, 2002, 21:38:03
Job time: 7092 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:32:46 ; Search time 76.51 seconds
(without alignments)
103.990 Million cell updates/sec

Title: US-09-391-606-7

Perfect score: 1166

Sequence: 1 MTKKHAWVVEGILNRLPKQ.....DKLGSDFTRFKDLGIISAF 217

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39;*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	92	7.9	667	TKT_BACSU	P45694 bacillus su
2	84	7.2	810	COAT_JCDNV	Q90053 junonia coe
3	81.5	7.0	449	EFIA_TRYBB	P41166 trypanosoma
4	80	6.9	225	MTRO_METKA	Q32864 methanopyru
5	80	6.9	1786	1 YCF1_ARATH	P56785 arabidopsis
6	77.5	6.6	250	1 VACJ_HAEIN	P44042 haemophilus
7	77.5	6.6	476	1 COAT_BACSD	Q929X0 bacillus ha
8	77.5	6.6	3801	1 LYST_HUMAN	Q99698 homo sapien
9	77	6.6	228	1 UCRI_SCHPO	Q09154 schizosacch
10	77	6.6	500	1 ABFA_BACSU	P94531 bacillus su
11	77	6.6	1109	1 CSAA_BACTF	Q9x682 bacillus th
12	76.5	6.6	755	1 DNLI_YEAST	P04819 saccharomyc
13	76	6.5	811	1 COAT_GMDNV	Q90125 galliera me
14	75.5	6.5	431	1 NORE_CHLPN	Q92723 chlamydia p
15	75	6.4	542	1 ESTS_DROVI	Q05487 drosophila
16	75	6.4	789	1 LFIA_DROME	P11995 drosophila
17	75	6.4	1311	1 STRH_STRPN	P49610 streptococc
18	74.5	6.4	825	1 EGLS_HANAN	P06835 hansenua a
19	74	6.3	700	1 NONA_DROME	Q04047 drosophila
20	74	6.3	789	1 LP1B_DROME	P11996 drosophila
21	73.5	6.3	315	1 OMPP_ECOLI	P34210 escherichia
22	73.5	6.3	445	1 OST4_CAEEL	P45971 caenorhabdi
23	73.5	6.3	456	1 DNAL_CHLTR	Q84252 chlamydia t
24	73.5	6.3	752	1 PSAA_ODOSI	P49479 odontella s
25	73.5	6.3	988	1 TNP6_ENTFC	Q06238 enterococcu
26	73	6.3	530	1 YA9A_SCHPO	Q09788 schizosacch
27	73	6.3	591	1 YF9A_SCHPO	Q09788 schizosacch
28	72.5	6.2	283	1 SOHA_RAT	P22789 rattus norv
29	72.5	6.2	291	1 THIG_CYACA	Q19915 cyanidium c
30	72.5	6.2	557	1 CO9_RABIT	P48747 cryptotolagu
31	72.5	6.2	572	1 PTLB_STAAU	P11162 staphylococ
32	72.5	6.2	809	1 COAT_DSDNV	Q71155 diatraea sa
33	72.5	6.2	931	1 TRP6_HUMAN	Q9y210 homo sapien

RESULT 1

ID	TKT_BACSU	STANDARD;	PRT;	667 AA.
AC	P45694;			
DT	01-NOV-1995 (Rel. 32, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DE	TRANSETOLASE (EC 2.2.1.1).			
GN	TKT OR TKT4.			
OS	Bacillus subtilis.			
OC	Bacteria; Firmicutes; Bacillus/Clostridium group;			
OC	Bacillus/Staphylococcus group; Bacillus.			
OX	NCBI_TaxID=1423;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=168;			
RA	Rose M., Entian K.;			
RL	Submitted (JUN-1996) to the EMBL/GenBank/DBJ databases.			
RN	[2]			
RP	SEQUENCE OF 561-667 FROM N.A.			
RC	STRAIN=168;			
RA	Schiott T., von Wachenfeldt C., Hederstedt L.;			
RL	Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.			
CC	-1- CATALYTIC ACTIVITY: SEDOPHTULOSE 7-PHOSPHATE + D-XYULOSE 5-PHOSPHATE.			
CC	3-PHOSPHATE - D-RIBOSE 5-PHOSPHATE + D-XYULOSE 5-PHOSPHATE.			
CC	1- COFACTOR: THIAMINE PYROPHOSPHATE.			
CC	1- SIMILARITY: BELONGS TO THE TRANSETOLASE FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announcement/ or send an email to license@isb-sib.ch).			
DR	EMBL; Z73234; CA97616.1; -			
DR	EMBL; X87845; CA61113.1; -			
DR	EMBL; Z99113; CAB13673.1; -			
DR	HSP; P23254; IAYO.			
DR	Subtilist; BG11247; tkt.			
DR	InterPro; IPR000360; Transketolase.			
DR	Pfam; PF00455; transketolase.1.			
DR	PROSITE; PS00801; TRANSETOLASE_1; 1.			
DR	PROSITE; PS00802; TRANSETOLASE_2; 1.			
KW	transferase; Thiamine pyrophosphate; Complete proteome.			
SQ	SEQUENCE 667 AA; 72344 MW; D93BCAQD246148AF CRC64;			

Query Match 7.9%; Score 92; DB 1; Length 667;

Best Local Similarity 22.9%; Pred. No. 0.57;

Matches 50; Conservative 32; Mismatches 68; Indels 68; Gaps 12;

QY 1 MTKKHAWVVEGILNRLPKQFFVKCSVVVDWNTFVPSFTSTTEKAATNAMYKY-----C 54

DB 273 LTRKAYAWTYE-----EDFIVFSEVYEHFAVAVKESGKKEQWNAQFAKYEVYPPEL 325

QY 55 VQWLVGKHQSVWINGOKKPLYLYGAFMLNPLAKATKTTLNG-KENLAWFIGGT--LGG 111
DB 326 AEQLELAITKGLPKDQDQVYVEKGGSLAS--RASSEVLNGLAKKIPFVGGADLAG 383
QY 112 -----LRKAGWSATRYEYVEALSVPEDVSGIGRGNLKFV-----FAQAIANYDPK 161
DB 384 SNKTTIKNAGDFTA-----VDYSG-----KNWFGVREFAMGAALN---- 419
QY 162 EANSFTNYKFGS---ALYMYGITSLSFRAYGAYSKPA 196
DB 420 -----GMLHGLRVLFGGT-----FFVFSYDLRPA 444

RESULT 2

ID COAT_JCDNV STANDARD; PRT; 810 AA.
AC Q90053;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE COAT PROTEIN VP1 (STRUCTURAL PROTEIN VP1) [CONTAINS: COAT PROTEIN VP
DE (STRUCTURAL PROTEIN VP2); COAT PROTEIN VP3 (STRUCTURAL PROTEIN VP3);
DE COAT PROTEIN VP4 (STRUCTURAL PROTEIN VP4)].
GN VP.
OS Junonia coenia densovirus (JCDNV).
OC Viruses; ssDNA viruses; Parvoviridae; Densovirinae; Densovirus.
OX NCBI_TaxID=12524;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93033112; PubMed=1413502;
RA Dumas B., Jourdan M., Pascaud A.M., Bergoin M.;
RT "Complete nucleotide sequence of the cloned infectious genome of
RT Junonia coenia densovirus reveals an organization unique among
RT parvoviruses";
RL Virology 191:202-222(1992).
CC -1- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC OF A COMBINATION OF VP1, VP3, VP2 AND VP1 (BY SIMILARITY).
CC -1- ALTERNATIVE PRODUCTS: THE FOUR DIFFERENT COAT PROTEINS ARE
CC PRODUCED BY ALTERNATIVE INITIATION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; S47266; AAB23698.1; -
CC InterPro; IPR003433; denso_VP4.
CC Pfam; PF02336; denso_VP4; 1.
CC Coats protein; Alternative initiation.
KW CHAIN 1 810 COAT PROTEIN VP1.
FT CHAIN 278 810 COAT PROTEIN VP2.
FT CHAIN 323 810 COAT PROTEIN VP3.
FT CHAIN 374 810 COAT PROTEIN VP4.
FT CHAIN 810 AA; 87895 MW; 0D618F04ADD04DE4 CRC64;
SQ SEQUENCE 810 AA; 87895 MW; 0D618F04ADD04DE4 CRC64;

Query Match 7.28; Score 84; DB 1; Length 810;
Best Local Similarity 25.8%; Pred. No. 3.9;
Matches 61; Conservative 22; Mismatches 59; Indels 94; Gaps 18;

QY 2 TKKHYAVWVGTLRLPKQFFVKCSVDWNTFPSESTTEKAATNAMKYKVCWQMLVG 61
DB 413 TKSH-KFMIFGLN-----NVIGTIGTGT---AVNRL-ITLCL----- 446
QY 62 KHSQVPTWINGOKKPLYLYGA-----FLMPLA-----KATK-TTLN 96
DB 447 --AEIPW--OKLPLMYNQSEFDLLPGSRVVECNVVKVIFRTNRIAPETSTATKQATLN 501
QY 97 KENLAWFIGTGLGLRAGDWSATRYEYVEALSVPEDVSGIGRGNLKFWEAQAI-- 154

DB 502 QISNLQ-----TAVGLNKLKLG-W-----GIDR-SFTAFQSDQPMIP 534
QY 155 -----AANYDKPEANSFTNYKGFSAFYMYGITSLSFRAYGAYSKPANDKIGSDPTF 206
DB 535 TATSAPKYEY--ITGTTGYRGMIAFY-YG-ADSTNDAAFGNAGNYPHQVGS-PTF 585

RESULT 3

ID EFIA_TRYBB STANDARD; PRT; 449 AA.
AC P41166;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE ELONGATION FACTOR 1-ALPHA (EF-1-ALPHA).
GN TEFL.
OS Trypanosoma brucei brucei.
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OX NCBI_TaxID=5702;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=LVH/75/USAMRU-K/18;
RX MEDLINE=94364999; PubMed=8083206;
RA Kaur K.J., Ruben L.;
RT "Protein translation elongation factor-1 alpha from Trypanosoma
RT brucei binds calmodulin";
RL J. Biol. Chem. 269:23045-23050(1994).
RN [2]
RP SEQUENCE OF 18-415 FROM N.A.
RX MEDLINE=94089672; PubMed=8265589;
RA Baldauf S.L., Palmer J.D.;
RT "Animals and fungi are each other's closest relatives: congruent
RT evidence from multiple proteins";
RL Proc. Natl. Acad. Sci. U.S.A. 90:11558-11562(1993).
CC -1- FUNCTION: THIS PROTEIN PROMOTES THE GTP-DEPENDENT BINDING OF
CC AMINOACYL-TRNA TO THE A-SITE OF RIBOSOMES DURING PROTEIN
CC BIOSYNTHESIS.
CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
CC -1- SIMILARITY: BELONGS TO THE GTP-BINDING ELONGATION FACTOR FAMILY.
CC EF-TU/EF-1A SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L25868; AAA16602.1; -
CC EMBL; U10562; AAA57476.1; -
CC HSP: Q01698; ITUI
DR InterPro; IPR000795; GTP_EFTU.
DR Pfam; PF00009; GTP_EFTU; 1.
DR PRINTS; PR00315; ELONGATNFCT.
DR PROSITE; PS00301; EFATOR_GTP; 1.
KW Elongation factor; Protein biosynthesis; GTP-binding.
FT NP_BIND 14 21 GTP (BY SIMILARITY).
FT NP_BIND 91 95 GTP (BY SIMILARITY).
FT NP_BIND 153 156 GTP (BY SIMILARITY).
FT CONFLICT 240 242 TCT -> DVY (IN REF. 2).
SQ SEQUENCE 449 AA; 49033 MW; F21113FB80A5116A CRC64;

Query Match 7.08; Score 81.5; DB 1; Length 449;
Best Local Similarity 21.5%; Pred. No. 3.5;
Matches 42; Conservative 25; Mismatches 71; Indels 57; Gaps 9;

QY 19 KQFFVKCSVD-----WNTFVPSSTTEKAATNAMKYKVC---WQW-LYKGH 63
DB 146 QMWVCCNKMKDDKTVNYQERYDEIVKESAYIKKGVNVEKRVFVPISGQGNMIEKS 205

QY	86	PLAKATKTLNGKENLAW	-----FIGTGLGGLKRGKGSWSTVRYEYVEALS	131	
QY	87	PLAKATKTLNGKENLAW	-----FIGTGLGGLKRGKGSWSTVRYEYVEALS	131	
Db	102	CSAKVQDKPTIGWDQEAIVTPGTEGHGIPTVSFVSGIYKGLL	LGSSGGAMVYIYALYKVLG	160	
QY	132	VPSIDVSGIGRGNLLKFWFAQAIAANY	-----DPKEANSFTNY-KGFSALYMYGI	180	
Db	161	MSAALAGILAMG----	FFVANAVLASYNIGGTTIEGYHDPK----	FTRLPKAVVCSLVFGI	212
QY	181	TDSL	184		
Db	213	VASV	216		

```

RESULT 5
ID YCFL_ARATH STANDARD; PRT; 1786 AA.
AC P56785;
DC AT
DT 30-MAY-2000 (Rel. 39, Created)
DD 30-MAY-2000 (Rel. 39, Last sequence update)
DE 20-AUG-2001 (Rel. 40, Last annotation update)
DN HYPOTHETICAL 213.7 KDA PROTEIN YCFL.
GS YCFL.
GN Arabidopsis thaliana (Mouse-ear cross).
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxId:3702;
[1]
RN SEQUENCE FROM N.A.
RP STRAIN=CV. COLUMBIA;
RX MEDLINE=20039611; PubMed=10574454;
RA Sato S., Nakamura Y., Kaneko T., Asamizu E., Tabata S.:
RT "Complete structure of the chloroplast genome of Arabidopsis
RT thaliana.";
RL DNA Res. 6:283-290(1999).
RC -1- FUNCTION: NOT YET KNOWN.
CC -1- SIMILARITY: BELONGS TO THE YCFL FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collabora-
CC between the Swiss Institute of Bioinformatics and the EMBL outstat-
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no
CC modified and this statement is not removed. Usage by and for commer-
CC entities requires a license agreement (See http://www.isb-sib.ch/anno-
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AP000423; BAA84445.1; -
CC DR EMBL; AP000423; BAA84433.1; -
CC KW Chloroplast; Hypothetical protein.
SQ SEQUENCE 1786 AA; 213727 MW; CFFD2A4D776D7E5D CRC64;
-----
Query Match 6.9%; Score 80; DB 1; Length 1786;
Best Local Similarity 19.9%; Pred. No. 21;
Matches 33; Conservative 37; Mismatches 74; Indels 22; Gaps
QY 8 WVVEGILNRLPKOFFVKCSVVDW--NTFFVPSETSTTEKAATNAMKYKC---VWQWLVK 62
| : : : : : : : : : : : : : : : : : : : : : : : :
Db 840 WLDTGQIQIKLFFPYLK----PWHKSKFQASQKARLUKTKDKGEKNDFCLTAV---GM 891
| : : : : : : : : : : : : : : : : : : : : : : : :
QY 63 HSOVPWINGOKPLYLYGAFLLMPLAKAYKTKTLNGENLAWFIIGTGLGLRAGDSATV 122
| : : : : : : : : : : : : : : : : : : : : : : : :

```

```
QY      123 RYYVEALSVPEIDVSGIGRGNLLKWFQAIAANYDPKEANSFTN 168
          |   :   :   :   :   :   :   :   :   :   :   :   :
Db      943 AKETKNWLKNFIFIKISKRNLIPFGPREIYELNEPKKDSIISN 988
```

RESULT 6
VACJ HAEIN

```
AC P44042;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE VACJ LIPOPROTEIN HOMOLOG PRECURSOR.
GN VACJ OR H10718.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus
OC NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Eritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae RD.";
RL Science 269:496-512(1995).
CC -1- SUBCELLULAR LOCATION: ATTACHED TO THE OUTER MEMBRANE BY A LIPID
CC ANCHOR (BY SIMILARITY).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U32754; AAC22375.1;
DR TIGR; H10718;
DR PROSITE; PS00013; PROKAR_LIPOPROTEIN; 1.
KW Outer membrane; Lipoprotein; Signal; Complete proteome.
FT SIGNAL 1 18
FT CHAIN 19 250 VACJ LIPOPROTEIN HOMOLOG.
FT LIPID 19 19 N-ACYL DIGLYCERIDE (POTENTIAL).
SQ SEQUENCE 250 AA; 28110 MW; 7DDEC4FC587091BB CRC64;

Query Match 6.68; Score 77.5; DB 1; Length 250;
Best Local Similarity 29.68; Pred. No. 4.3;
Matches 32; Conservative 10; Mismatches 35; Indels 31; Gaps 6;

QY 28 VDMNTFV--PSETSTTEKAATNAM---KYKCYVMQWLVGKHSQVFWINGQKKPLYLYGAF 82
DB 144 VDAGTYIVLPYNATTPQLTCAVDAAAYMYPFQWVG---PW-----ALVKYGVQ 192

QY 83 LNNPLAKATKTLNKENLAWFIGTGLGLRKAGDWSATVRYVEAL 130
DB 193 AVDARAK----NLNNAEL-----LRQAQDPYITFREAYYQNL 225

RESULT 7
ID GATB_BACHD STANDARD; PRT; 476 AA.
AC Q949X0; Q9JPV7;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE GLUTAMYL-TRNA(GLN) AMIDOTRANSFERASE SUBUNIT B (EC 6.3.5.-) (GLU-ADT
DE SUBUNIT B).
GN GATB OR BH0667.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/staphylococcus group; Bacillus.
```

```
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=99184646; PubMed=10086842;
RA Takami H., Nakasone K., Ogasawara N., Hirama C., Nakamura Y.,
RA Masui N., Fuji F., Takaki Y., Inoue A., Horikoshi K.;
RT "Sequencing of three lambda clones from the genome of alkaliphilic
RT Bacillus sp. strain C-125.";
RT Extremophiles 3:29-34(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RX MEDLINE=20512582; PubMed=11058132;
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
RA Horikoshi K.;
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
RT halodurans and genomic sequence comparison with Bacillus subtilis.";
RT Nucleic Acids Res. 28:4317-4331(2000).
CC -1- FUNCTION: FURNISHES A MEANS FOR FORMATION OF CORRECTLY CHARGED
CC GLN-TRNA(GLN) THROUGH THE TRANSMIDATION OF MISACYLATED GLU-
CC TRNA(GLN) IN ORGANISMS WHICH LACK GLUTAMINYL-TRNA SYNTHETASE. THE
CC REACTION TAKES PLACE IN THE PRESENCE OF GLUTAMINE AND ATP THROUGH
CC AN ACTIVATED GAMMA-PHOSPHO-GLU-TRNA(GLN) (BY SIMILARITY).
CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMYL-TRNA(GLN) + L-GLUTAMINE = ADP
CC + PHOSPHATE + L-GLUTAMINYL-TRNA(GLN) + L-GLUTAMATE.
CC -1- SUBUNIT: HETEROTRIMER OF A, B AND C SUBUNITS (BY SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE GATB FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AB011836; BAA75312.1;
DR EMBL; AP001509; BAB04386.1;
DR InterPro; IPR001773; Gln_amidotransf_B.
DR Pfam; PF01162; PET112; 1.
DR PROSITE; PS01234; GATB; 1.
KW Protein biosynthesis; Ligase; Complete proteome.
SQ SEQUENCE 476 AA; 53414 MW; C212B522413C579B CRC64;

Query Match 6.68; Score 77.5; DB 1; Length 476;
Best Local Similarity 27.28; Pred. No. 8.6;
Matches 25; Conservative 13; Mismatches 39; Indels 15; Gaps 2;

QY 90 ATKTTL----NGKENLAWFIGTGLGLRKAGDWSATVRYE-----YVEALSYPE 134
DB 253 ANKTVLMRVKSGDDYRYFPEPDVLVALHIDDEWKARISEIPELDPARKKRYVEELGLPA 312

QY 135 IDVSGIGRGNLLKFWFAQAIANYDPKEANSF 166
DB 313 YDAMVLTITLKEMSDFFETIAKAGADPKLASNW 344

RESULT 8
ID LYST_HUMAN STANDARD; PRT; 3801 AA.
AC Q99698; Q99709; O43274;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE LYSOSOMAL TRAFFICKING REGULATOR (BEIGE HOMOLOG).
GN CHS1 OR LYST OR CHS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
```


[1]
 SEQUENCE FROM N.A. (ISOFORM 1).
 MEDLINE=97051925; PubMed=8896560;
 Nagle D.L., Karim M.A., Woolf E.A., Holmgren L., Bork P., Misumi D.J.,
 McGrail S.H., Dussault B.J., Perou C.M., Bolissy R.E., Duyk G.M.,
 Spritz R.A., Moore K.J.;
 RT Chediak-Higashi syndrome and mutation analysis of the complete gene for
 Chediak-Higashi syndrome.;
 Nat. Genet. 14:307-311(1996).
 [2]
 SEQUENCE FROM N.A. (ISOFORM 2).
 TISSUE=Liver;
 MEDLINE=96353977; PubMed=8717042;
 Barbosa M.D.F.S., Nguyen Q.A., Tcherev V.T., Ashley J.A.,
 Better J.C., Blaydes S.M., Brandt S.J., Chotai D., Hodgman C.,
 Solari R.C.E.S., Lovett M., Kingsmore S.F.;
 RT "Identification of the homologous beige and Chediak-Higashi syndrome
 genes";
 Nature 382:262-265(1996).
 [3]
 SEQUENCE FROM N.A. (ISOFORM 3).
 MEDLINE=97358584; PubMed=9215680;
 Barbosa M.D.F.S., Barrat F.J., Tcherev V.T., Nguyen Q.A.,
 Mishra V.S., Colman S.D., Pastural E., Dufourcq-Lagelouse R.,
 Fischer A., Holcombe R.F., Wallace M.R., Brandt S.J.,
 de Saint Basile G., Kingsmore S.F.;
 RT "Identification of mutations in two major mRNA isoforms of the
 Chediak-Higashi syndrome gene in human and mouse";
 Hum. Mol. Genet. 5:1091-1098(1997).
 [4]
 SEQUENCE OF 1-1475 FROM N.A.
 Barrat F.J., Barbosa M.D.F.S., Pastural E., Dufourcq-Lagelouse R.,
 Kingsmore S.F., de Saint Basile G.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: MAY REGULATE LYSOSOME TRANSPORT.
 CC -!- SUBCELLULAR LOCATION: CYTOPLASMIC (POTENTIAL).
 CC -!- ALTERNATIVE PRODUCTS: AT LEAST 3 ISOFORMS; 1 (SHOWN HERE), 2
 AND 3; ARE PRODUCED BY ALTERNATIVE SPLICING.
 CC -!- TISSUE SPECIFICITY: ABUNDANTLY EXPRESSED IN ADULT AND FETAL
 THYMUS, PERIPHERAL BLOOD LEUKOCYTES, BONE MARROW AND SEVERAL
 REGIONS OF THE ADULT BRAIN.
 CC -!- DISEASE: DEFECTS IN CHS1 ARE THE CAUSE OF CHEDIAK-HIGASHI SYNDROME
 (CHS) A RARE AUTOSOMAL RECESSIVE DISORDER CHARACTERIZED BY
 HYPOPIGMENTATION, SEVERE IMMUNOLOGIC DEFICIENCY, A BLEEDING
 TENDENCY, NEUROLOGIC ABNORMALITIES, ABNORMAL INTRACELLULAR
 TRANSPORT TO AND FROM THE LYSOSOME, AND GIANT INCLUSION BODIES IN
 A VARIETY OF CELL TYPES.
 CC -!- SIMILARITY: CONTAINS 7 WD REPEATS (TRP-ASP DOMAINS).
 CC -!- SIMILARITY: CONTAINS 1 BEACH DOMAIN.

 This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

 EMBL; U84744; AAB87737.1; .
 DR EMBL; U67615; AAB41309.1; .
 DR EMBL; U72192; AAB39697.1; .
 DR EMBL; L77889; AAB51608.1; .
 DR EMBL; U70064; AAB41533.1; .
 DR MIM; 214500; .
 DR InterPro; IPR000409; Beige_BEACH.
 DR InterPro; IPR001680; WD40.
 DR Pfam; PF02138; Beach; 1.
 DR Pfam; PF00400; WD40; 5.
 DR ProDom; PD007848; Beige_BEACH; 1.
 DR SMART; SM00320; WD40; 4.
 DR PROSITE; PS0197; BEACH; 1.
 DR PROSITE; PS00678; WD_REPEATS_1; 1.
 DR PROSITE; PS50082; WD_REPEATS_2; 1.

DR PROSITE; PS50294; WD_REPEATS_REGION; 1.
 KW Protein transport; Transport; Repeat; WD repeat; Alternative splicing.
 FT DOMAIN 31 36 POLY-GLU.
 FT DOMAIN 2448 2452 POLY-LEU.
 FT REPEAT 662 700 WD 1.
 FT REPEAT 1582 1626 WD 2.
 FT DOMAIN 3139 3422 BEACH.
 FT REPEAT 3563 3602 WD 3.
 FT REPEAT 3614 3653 WD 4.
 FT REPEAT 3656 3699 WD 5.
 FT REPEAT 3700 3744 WD 6.
 FT REPEAT 3749 3788 WD 7.
 FT VARSPLIC 1515 1531 ESDRPEGAEYINPGERL -> GMMTGLSDLYTKIVFRL
 (IN ISOFORM 3).
 FT VARSPLIC 1532 3801 MISSING (IN ISOFORM 3).
 FT VARSPLIC 1988 2001 VCRSFVKIIAEVLG -> MARSFRKCCQSCT (IN
 ISOFORM 2).
 FT VARSPLIC 2002 3801 MISSING (IN ISOFORM 2).
 SQ SEQUENCE 3801 AA; 429117 MW; 288B370AD4BDAF11 CRC64;

 Query Match 6.6%; Score 77.5; DB 1; Length 3801;
 Best Local Similarity 22.2%; Pred.No. 82;
 Matches 41; Conservative 15; Mismatches 50; Indels 79; Gaps 7;
 QY 31 NTFVPSETSTTEKAATNAMYKYCYVQWLVGKHQSV-----WINGQKKPLYL 78
 Db 1585 NIEFSPK-----WQHLVLYLQOPQCKRRIHKGISIWVSQKRPDVT 1626
 QY 79 YGAFLMPLAKATKTLNGKENLAW-FITGTLGG-----LRKAGDWSATVRYEVEALSVP 133
 Db 1627 LDFML-----PRKTSLSDSNKTFCMIGHCLSSOEFLQLAGKWDL----- 1667
 QY 134 EIDVSGICRGNLLKFEWAQIAAAYNDPREANSFTNYKGFSAALYMYGITDLSLFRAYGAYS 193
 Db 1668 -----GNLLFNGAKVGSQ-----EAYLYACGPNHTSVMPKYG 1702
 QY 194 KPAND 198
 Db 1703 KPVND 1707

 RESULT 9
 UCRL SCHPO
 ID UCRL SCHPO STANDARD; PRT; 228 AA.
 AC Q09154; O42942;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE UBIOQUINOL-CYTOCHROME C REDUCTASE IRON-SULFUR SUBUNIT, MITOCHONDRIAL
 PRECURSOR (EC 1.10.2.2) (RIESKE IRON-SULFUR PROTEIN) (RISP).
 GN RIPI OR SPBC16H5.06.
 OS Schizosaccharomyces pombe (Fission yeast).
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
 OC Schizosaccharomycetales; Schizosaccharomycetaceae;
 OC Schizosaccharomycetes.
 OX NCBI_TaxID=4896;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=96279043; PubMed=8663290;
 RA Diago J., Bruel C., Graham L.A., Slonimski P., Trumppower B.L.;
 RT "Heterologous complementation of a Rieske iron-sulfur protein-
 deficient Saccharomyces cerevisiae by the ripl gene of
 Schizosaccharomyces pombe";
 RL J. Biol. Chem. 271:15341-15345(1996).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=972;
 RA Wood V., Rajandream M.A., Barrell B.G., Skelton J., Churcher C.M.;
 RL Submitted (MAR-1998) to the EMBL/GenBank/DBJ databases.
 CC -!- FUNCTION: COMPONENT OF THE UBIOQUINOL-CYTOCHROME C REDUCTASE
 COMPLEX (COMPLEX III OR CYTOCHROME B-C1 COMPLEX), WHICH IS A
 RESPIRATORY CHAIN THAT GENERATES AN ELECTROCHEMICAL POTENTIAL

```

CC COUPLED TO ATP SYNTHESIS.
CC -!- CATALYTIC ACTIVITY: QH(2) + 2 FERRICYTOCHROME C = Q +
CC 2 FERROCYTOCHROME C.
CC -!- SUBUNIT: FUNGI BCL COMPLEX CONTAINS 10 SUBUNITS; 3 RESPIRATORY
CC SUBUNIT, 2 CORE PROTEINS AND 5 LOW-MOLECULAR WEIGHT PROTEINS.
CC -!- SUBCELLULAR LOCATION: MITOCHONDRIAL INNER MEMBRANE.
CC -!- MISCELLANEOUS: THE RIESKE PROTEIN IS A HIGH POTENTIAL 2FE-2S
CC PROTEIN.
CC -!- SIMILARITY: TO RIESKE PROTEINS FROM OTHER SOURCES (MITOCHONDRIA,
CC BACTERIAL, CHLOROPLAST).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U04080; AAC49359.1; -.
CC EMBL; AL022104; CAAL7904.1; -.
CC HSSP; P13272; IRIE.
CC InterPro; IPR001700; RNA_pol_A_bac.
CC InterPro; IPR001281; Rieske.
CC Pfam; PF00355; Rieske; 1.
CC PRINTS; PR00162; RIESKE.
CC ProDom; PD001179; RNA_pol_A_bac; 1.
CC PROSITE; PS00199; RIESKE_1; 1.
CC PROSITE; PS00200; RIESKE_2; 1.
CC Mitochondrion; Electron transport; Respiratory chain; Iron-sulfur;
CC Oxidoreductase; Inner membrane; Transmembrane; Transit peptide.
CC TRANSIT 1 ? MITOCHONDRION (POTENTIAL).
CC CHAIN ? 228 UBIQUINOL-CYTOCHROME C REDUCTASE IRON-
CC SULFUR SUBUNIT.
CC METAL 172 172 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 174 174 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 191 191 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC METAL 194 194 IRON-SULFUR (2FE-2S) (BY SIMILARITY).
CC DISULFID 177 193 BY SIMILARITY.
CC CONFLICT 183 183 A -> R (IN REF. 1).
CC SEQUENCE 228 AA; 24740 MW; C50CC8BA159E31B2 CRC64;

Query Match 6.6%; Score 77; DB 1; Length 228;
Best Local Similarity 30.0%; Pred. No. 4.3;
Matches 27; Conservative 17; Mismatches 28; Indels 18; Gaps 4;

QY 91 TKTTLNGENLAWFGTGLGLRKGKAG-----DSSATVRYEYVEALSVPEDVSGI 140
DB 58 TKTSGDRSRVISYAVGTMGALTAAGAAQATVHDFLASWSASAD---VLAMSKAEVDLSKI 114
QY 141 GRG-NLLKFWFAQAIAANY----DPKEANS 165
DB 115 PEGKNLVVKWQCKPFIHRPPEEQEANS 144

RESULT 10
ABFA_BACSU STANDARD; PRT; 500 AA.
AC P94531; O05096;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE ALPHA-L-ARABINOFURANOSIDASE 1 (EC 3.2.1.55) (ARABINOSIDASE).
GN ABFA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97124191; PubMed=8969504;

```

```

RA Wipat A., Carter N., Brignell C.S., Guy J.B., Piper K.,
RA Sanders J., Emmerson P.T., Harwood C.R.;
RT "The dnaB-phage (256 degrees-240 degrees) region of the Bacillus
RT subtilis chromosome containing genes responsible for stress
RT responses, the utilization of plant cell walls and primary
RT metabolism.";
RL Microbiology 142:3067-3078(1996).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=97237725; PubMed=9084180;
RA Sa-Nogueira I., Nogueira T.V., Soares S., de Lencastre H.;
RT "The Bacillus subtilis L-arabinose (ara) operon: nucleotide sequence,
RT genetic organization and expression.";
RL Microbiology 143:957-969(1997).
CC -!- CATALYTIC ACTIVITY: HYDROLYSIS OF TERMINAL NON-REDUCING ALPHA-L-
CC ARABINOFURANOSIDE RESIDUES IN ALPHA-L-ARABINOSIDES.
CC -!- SIMILARITY: BELONGS TO FAMILY 51 OF GLYCOSYL HYDROLASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; Z75208; CAA9595.1; -.
CC EMBL; X89810; CAA61937.1; -.
CC EMBL; Z99118; CAB14832.1; -.
CC Subtilist; Bg11900; abfa.
CC Hydrolase; Glycosidase; Complete proteome.
CC CONFLICT 114 114 P -> A (IN REF. 2).
CC SEQUENCE 500 AA; 57060 MW; 7397ED52A4987686 CRC64;

Query Match 6.6%; Score 77; DB 1; Length 500;
Best Local Similarity 23.1%; Pred. No. 10;
Matches 50; Conservative 33; Mismatches 89; Indels 44; Gaps 13;

QY 15 NLRPKQFFVKCSVDWNTFVPSSTTE--KAATNAMK-YKYCVWQVLVGRHSQ----- 65
DB 250 NNLPNYL---ARSMDLDFIKSVAATCDYVRAKTRSKRTINLSLDENWVYHSNEADKKV 306
QY 66 VPWLNGOK--KPLY-----LYGAFLMNPLAKATKTLNGENLAWFGTGLGLRKGKAG 117
DB 307 EPWITAREILEDIYNFEDALLVGSLLITMLQHADRKVKTAQLQVNVNVIAPIM--TEKGE 364
QY 118 -WSATVRYEYVEALSVPEDVSGIGRGNLLKFWFAQAIAANYDPK-EANSFTNYKGFSA 175
DB 365 AWRQPIFYPMHA-----SVYGRGSLKPLIS-----SPKYDCSDFTDVPYVDAA 409
QY 176 YMYG-ITDSLSFRAYGAYSKPANDKLGSDFTFRKFD 210
DB 410 VVYSEETLTIFAV---NKAEDQMETEISLRGFE 441

RESULT 11
CSAA_BACTF STANDARD; PRT; 1109 AA.
AC Q9X682;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PESTICIDIAL CRYSTAL PROTEIN CRY28AA (INSECTICIDAL DELTA-ENDOTOXIN
DE CRYXXVIII(A)) (CRYSTALLINE ENTOMOCIDAL PROTOXIN) (126 KDA CRYSTAL
DE PROTEIN).
GN CRY28AA OR CRYXXVIII(A).
OS Bacillus thuringiensis (subsp. finitimus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=29337;
RN [1]

```

SEQUENCE FROM N.A.
 RT STRAIN-VKPM B-1161;
 RC MEDLINE=99330166; PubMed=10403372;
 RA Wojciechowska J.A., Lewitin E., Revina L.P., Zalunin I.A.,
 RA Chestukhina G.G.;
 RT "Two novel delta-endotoxin gene families cry26 and cry28 from *Bacillus thuringiensis* ssp. *finitimus*.";
 RL FEBS Lett. 453:46-48(1999).
 CC !- FUNCTION: PROMOTES COLLOIDSMOTIC LYSIS BY BINDING TO THE MIDGUT
 CC !- DEVELOPMENTAL STAGE: THE CRYSTAL PROTEIN IS PRODUCED DURING
 CC !- OF THE SPORE COAT.
 CC !- MISCELLANEOUS: TOXIC SEGMENT OF THE PROTEIN IS LOCATED IN THE
 CC N-TERMINUS.
 CC !- SIMILARITY: BELONGS TO THE DELTA ENDOTOXIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AF132928; AAD24189.1; -;
 DR InterPro; IPR001178; Endotoxin.
 DR Pfam; PF00555; endotoxin; 1.
 KW Toxin; Sporulation.
 SQ SEQUENCE 1109 AA; 125712 MW; 10C80705508F5CDA CRC64;

 Query Match 6.6%; Score 77; DB 1; Length 1109;
 Best Local Similarity 27.0%; Pred. No. 24;
 Matches 51; Conservative 23; Mismatches 67; Indels 48; Gaps 12;
 QY 15 NRLPKOFFVKCSVDNWFVPSST--TEKAATNAMYKVCV--WQMLVGK-----H 63
 DB 387 NEYKFNFNVN-----NONEPQETTPNDYGGNSQKFKHLSHFLLIHKLEFAEYFH 440
 QY 64 S--QVPMW---INGOKKPLYLGAFLMPLAKATKTLNGKLENLAWFTGGTGLGLR--- 113
 DB 441 SIFALGWTHNSVNSN--LISESVSTQPLVKAYEVTNNSVIRGPGFTGGDLIEURDKCS 498
 QY 114 ---KAGD---WSATVRYEVEALSPEIDVSGIGRGNLLKFWFAQAIAYNDPKANSFT 167
 DB 499 IKCKASSLLKYAISLFYANNAIAV-SIDVGDGAGVLLQPTFSR-----KGNNT 549
 QY 168 ----NYKGF 172
 DB 550 IODLNYKDF 558

 RESULT 12
 DNLI_YEAST
 ID DNLI_YEAST STANDARD; PRT; 755 AA.
 AC P04819; Q12736;
 DT 13-AUG-1987 (Rel. 05, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE DNA LIGASE I, MITOCHONDRIAL PRECURSOR (EC 6.5.1.1)
 DE (POLYDEOXYRIBONUCLEOTIDE SYNTHASE [ATP]).
 GN CDC9 OR YDL164C.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
 OX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=86093646; PubMed=3909103;
 RA Barker D.G., White J.H.M., Johnston L.H.;
 RT "The nucleotide sequence of the DNA ligase gene (*CDC9*) from
 RT Saccharomyces cerevisiae: a gene which is cell-cycle regulated and

induced in response to DNA damage.";
 Nucleic Acids Res. 13:8323-8337(1985).
 [2]
 RN SEQUENCE FROM N.A.
 RP STRAIN=S288C;
 RC Pohl T.M.;
 RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
 [3]
 RP SEQUENCE OF 610-755 FROM N.A.
 RC STRAIN=AH22;
 RX MEDLINE=93247548; PubMed=8483449;
 RA Wehner E.P., Rao E.P., Brendel M.;
 RT "Molecular structure and genetic regulation of SFA, a gene
 RT responsible for resistance to formaldehyde in *Saccharomyces*
 RT *cerevisiae*, and characterization of its protein product.";
 RT Mol. Gen. Genet. 237:351-358(1993).
 RL [4]
 RP ALTERNATIVE INITIATION.
 RC MEDLINE=20003237; PubMed=10531002;
 RX Willer M., Rainey M., Pullen T., Stirling C.J.;
 RT "The yeast *CDC9* gene encodes both a nuclear and a mitochondrial form
 RT of DNA ligase I.";
 RL Curr. Biol. 9:1085-1094(1999).
 CC !- FUNCTION: THIS PROTEIN SEALS NICKS IN DOUBLE-STRANDED DNA DURING
 CC DNA REPLICATION, DNA RECOMBINATION AND DNA REPAIR. THE
 CC MITOCHONDRIAL FORM IS REQUIRED FOR MITOCHONDRIAL DNA MAINTENANCE
 CC BUT IS NON-ESSENTIAL WHILE THE NUCLEAR FORM IS ESSENTIAL FOR CELL
 CC VIABILITY.
 CC !- CATALYTIC ACTIVITY: ATP + (DEOXYRIBONUCLEOTIDE)(N) +
 CC (DEOXYRIBONUCLEOTIDE)(M) = AMP + PYROPHOSPHATE +
 CC (-) SUBCELLULAR LOCATION: NUCLEAR AND MITOCHONDRIAL.
 CC !- ALTERNATIVE PRODUCTS: A SINGLE GENE PRODUCES BOTH A MITOCHONDRIAL
 CC FORM AND A NUCLEAR FORM BY THE USE OF ALTERNATIVE INITIATION
 CC CODONS IN THE SAME READING FRAME.
 CC !- MISCELLANEOUS: CDC9 IS INCLUDED WITHIN THE CATEGORY OF SO-CALLED
 CC 'START GENES', ENCODING PROTEINS WHICH ARE REQUIRED IN EARLY G1,
 CC WHEN THE CELL IS FACED WITH THE OPTION OF INITIATING A FURTHER
 CC CELL CYCLE.
 CC !- SIMILARITY: BELONGS TO THE ATP-DEPENDENT DNA LIGASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03246; CAA27005.1; -;
 DR EMBL; 267750; CAA91582.1; -;
 DR EMBL; 274212; CAA98737.1; -;
 DR EMBL; X68020; CAA48158.1; -;
 DR PIR; S05830; LOBYPX.
 DR PIR; S31138; S31138.
 DR SGD; S0002323; CDC9.
 DR InterPro; IPR000977; DNA_ligase.
 DR Pfam; PF01068; DNA_ligase; 1.
 DR PROSITE; PS00697; DNA_LIGASE_A1; 1.
 DR PROSITE; PS00333; DNA_LIGASE_A2; 1.
 DR PROSITE; PS50160; DNA_LIGASE_A3; 1.
 KW DNA repair; DNA replication; DNA recombination; Cell division; Ligase;
 KW ATP-binding; Nuclear protein; Mitochondrion; Alternative initiation;
 KW Transit peptide.
 FT TRANSIT 1 44 MITOCHONDRION (POTENTIAL).
 FT CHAIN 45 755 DNA LIGASE I, MITOCHONDRIAL ISOFORM.
 FT CHAIN 24 755 DNA LIGASE I, NUCLEAR ISOFORM.
 FT INIT_MET 24 24 FOR NUCLEAR ISOFORM.
 FT BINDING 419 419 AMP (BY SIMILARITY).
 FT CONFLICT 69 69 D -> E (IN REF. 1).
 FT CONFLICT 186 186 L -> V (IN REF. 1).
 FT CONFLICT 671 671 G -> E (IN REF. 3).
 FT CONFLICT 724 724 R -> I (IN REF. 3).

SQ SEQUENCE 755 AA; 84828 MW; B7C2ECAF5C61CAE7 CRC64;
Query Match 6.6%; Score 76.5; DB 1; Length 755;
Best Local Similarity 21.08; Pred. No. 18; Indels 53; Gaps 7;
Matches 45; Conservative 26; Mismatches 90;
QY 2 TKKHAYVVEGILNRLP-----KQFFVKCVVDWN-----TFVPSETSTTEK 43
DB 434 TMRYSRNGENWTERPEINITDFIQDLDTTKNLILDEAVAWDKDQKILPFOVLSTRK 493
QY 44 ---AATNAMKYCYVQWMLVKHSHOVWINGOKKPLYLYGAFLMNPPLAKATKTTLNGKEN 100
DB 494 RKDELNDVKVKVCLFADFILCYNDERLINKSKERREY-----LTKVKVVPGEFQY 546
QY 101 LAWFITGTLGLRKAGDSATVRYE--YVEALSVPEDVSGIGRGNLLKFWFAQAIARA 158
DB 547 AFQITTNLDELQKFLDESVNHSCEGLMKVLEGE-----SHY 585
QY 159 DPKEANSFTNYKFSALYMYGITSLSFRAYGAY 192
DB 586 EPSKRS--RNWLKLLKDYLEGVDSLDLCVLGAY 617

RESULT 13
COAT_GMDNV STANDARD; PRT; 811 AA.
AC 090125; 090128; 090126; 090127;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DE COAT PROTEIN VP1 (STRUCTURAL PROTEIN VP1) [CONTAINS: COAT PROTEIN VP
DE (STRUCTURAL PROTEIN VP2); COAT PROTEIN VP3 (STRUCTURAL PROTEIN VP3);
DE COAT PROTEIN VP4 (STRUCTURAL PROTEIN VP4)].
GN VP1
OS Galleria mellonella densovirus (GMDNV).
OC Viruses; ssDNA viruses; Parvoviridae; Densovirinae; Densovirus.
OX NCBI_TaxID=37138;
RN [1]
RP SEQUENCE FROM N.A.
RA Tijssen P.;
RT "Organization and expression of the ambisense genome of
RT densoleucosis virus of Galleria mellonella (GMDNV).";
RL Submitted (NOV-1994) to the EMBL/GenBank/DBJ databases.
RN [2]
RP X-RAY CRYSTALLOGRAPHY (3.6 ANGSTROMS) OF 396-811.
RX MEDLINE=99036856; PubMed=9817847;
RA Simpson A.A., Chipman P.R., Baker T.S., Tijssen P., Rossmann M.G.;
RT "The structure of an insect parvovirus (Galleria mellonella
RT densovirus) at 3.7 A resolution.";
RL Structure 6:1355-1367(1998).
CC -|- SUBUNIT: THE VIRUS CAPSID IS COMPOSED OF 60 ICOSAEDRAL UNITS,
CC OF A COMBINATION OF VP4, VP3, VP2 AND VP1.
CC -|- ALTERNATIVE PRODUCTS: THE FOUR DIFFERENT COAT PROTEINS ARE
CC PRODUCED BY ALTERNATIVE INITIATION.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; L32896; AAA66966.1; -
CC DR EMBL; L32896; AAA66964.1; -
CC DR EMBL; L32896; AAA66965.1; -
CC DR EMBL; L32896; AAA66967.1; -
CC DR PDB; 1DNV; 16-FEB-99.
CC InterPro: IPR003433; denso.VP4.
CC Pfam: PF02336; denso.VP4; 1.
KW Coat protein; Alternative initiation; 3D-structure.
FT CHAIN 1 811 COAT PROTEIN VP1.

FT CHAIN 277 811 COAT PROTEIN VP2.
FT CHAIN 324 811 COAT PROTEIN VP3.
FT CHAIN 375 811 COAT PROTEIN VP4.
SQ SEQUENCE 811 AA; 88542 MW; 0A61B09172DF99B5 CRC64;
Query Match 6.5%; Score 76; DB 1; Length 811;
Best Local Similarity 22.4%; Pred. No. 21;
Matches 59; Conservative 36; Mismatches 82; Indels 86; Gaps 16;
QY 2 TKKHAYVVEGILNRLPKQFFVKCVVDWNTFFPSETSTTEKATNAMKYCYVQWMLVG 61
DB 414 TKSH-KFMIFLAN-----NVIGTGTGT---AVNRL-----LTT 445
QY 62 KHSOVWINGOKKPLYLYGA-----FLMNPPLAKAT-----KTTLN 96
DB 446 CLAEIPW---QKPLYNQSEFDLLPGRSVVECNVVFETNRIATSETSTVTKQATLN 502
QY 97 KENLAWFIQGTGLGRKAGDSATVRYEVE-----ALSVPEID-VSG-IG-RGNL 145
DB 503 QISNVQ-----TAIGLNKLG-WGINRAFTAFQSDQPMIPTATTAPKYEPVTGDTGYRCMI 556
QY 146 LKFWFAQAI-----AANYDPKEANSFTNYKFSALYMY-----GITDLSFRAYGAYS 193
DB 557 ADYTGADSTNDTAFGNAGNYPHHQVSSFTFLQNYCYMYQQTNGTGWPCLAHLQOFDS 616
QY 194 KPANDKLGSDFTFRKFDLGIISA 216
DB 617 KTVNQCLIDVY-KPKMGLIKS 638
RESULT 14
ID_NORF_CHLPN STANDARD; PRT; 431 AA.
AC 097723; 09J595;
DT 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DT 20-AUG-2001 (Rel. 40, Last annotation update)
DE PROBABLE NA(+)-TRANSLLOCATING NADH-QUINONE REDUCTASE SUBUNIT F
DE (EC 1.6.5.-) (NA(+)-TRANSLLOCATING NOR SUBUNIT F) (NA(+)-NOR SUBUNIT F)
DE (NOR COMPLEX SUBUNIT F) (NOR-1 SUBUNIT F).
GN NORF OR NOR6 OR CPN0883 OR CP0983
OS Chlamydia pneumoniae (Chlamydia pneumoniae).
OC Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia.
OX NCBI_TaxID=83558;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CWL029;
RX MEDLINE=99206606; PubMed=10192388;
RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
RA Olinger L., Grimwood J., Davis R.W., Stephens R.S.;
RT "Comparative genomes of Chlamydia pneumoniae and C. trachomatis.";
RL Nat. Genet. 21:385-389(1999).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=AR39;
RX MEDLINE=20150255; PubMed=10684935;
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
RA White O., Hickey E.K., Peterson J., Utterback T., Berry K., Bass S.,
RA Linher K., Weidman J., Khouri H., Craven B., Bowman C., Dodson R.,
RA Gwinn M., Nelson W., DeBoy R., Kolonay J., McClarty G., Salzberg S.L.,
RA Eisen J., Fraser C.M.;
RT "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
RT pneumoniae AR39.";
RL Nucleic Acids Res. 28:1397-1406(2000).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=J138;
RX MEDLINE=20330349; PubMed=10871362;
RA Shirai M., Hirakawa H., Kimoto M., Tabuchi M., Kishi F., Ouchi K.,
RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
RT "Comparison of whole genome sequences of Chlamydia pneumoniae J138
RT from Japan and CWL029 from USA.";

Db 320 GNFAQVPLASYTPENGICYNAALL--LAK-----DANGKBERIEELNTRWNELAPYFLAYPY 373
QY 108 TLGGLRKAGDWSA---TVRYEYVEALSVPEIDVSGIGRGNLLKFWFAQAIANYDPKEAN 164
Db 374 TL-----KRSENNAHSQKLYQYL-----GYKNFSVNYFDVQRL----- 408
QY 165 SFTNYKGFSAIYMWGITDSL-SFRAYGA-----YSKPANDKLGSDFTFRKFDLGI 213
Db 409 -FTN-----ELYKKGIELSLDSHRKHGASPVYAYVDNPA-DKSLAQFLAKRSDISL 458

Search completed: February 7, 2002, 21:42:39
Job time: 593 sec

PA (AVET) AVENTIS PASTEUR LTD.

```

XX  Murdin AD, Oomen RP, Wang J, Dunn P;
XX  WPI; 2000-687542/67.
XX  N-PSDB; AAD02066.
XX  Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX  useful for vaccinating against Chlamydia infections -
XX  Claim 33; Fig 3; 112pp; English.
XX  The present sequence is a fusion protein comprising a
XX  truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
XX  residues. C. pneumoniae 76 kDa protein is used in the
XX  diagnosis, prevention and treatment of C. pneumoniae
XX  infections (e.g. pneumonia, upper respiratory
XX  tract disease, bronchitis, sinusitis and acute respiratory
XX  disease such as cough, sore throat, hoarseness, fever; and
XX  abnormal chest sounds on auscultation). C. pneumoniae sequence
XX  is also used as vaccines for immunising humans against diseases
XX  caused by C. pneumoniae.
XX  Sequence 490 AA;

Query Match          99.2%; Score 2392; DB 21; Length 490;
Best Local Similarity 99.2%; Pred. No. 5.8e-153;
Matches 486; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY  1 MNPVIGPGPIDETERTPPADLSAAGLEASAAKSAEQAQRIAGAEAKPKESKTSVERWSI 60
DB  1 mnpvpgpiddetertppadlsaggleasaanksaeeqriagaekpkesktsverwsi 60
QY  61 LRSVAVNLSLADKLGITASSNSSSTSRSDVDSTTATPTPPPTSDYKTAQTAQYDT 120
DB  61 lrsavnlmsladvkltgiasnssstsrsvdvdsttatptppptdyktaqtaqydt 120
QY  121 IFTSTSLADIQAALVSLQDAVTNPKDTAAIDETAIAAEWETKNADAIKVGAQITELAKY 180
DB  121 iftstsladiqaalvslqdvavtnikdtaateetaiaaeawetknadvkvgaqitelaky 180
QY  181 ASDNQAILDLSGLKTSFDLLQTAALQSVANNKKAELLKEMQDNPVPGKTPAQAQSLVD 240
DB  181 asdnqaildlsigkltstfdllqtaalqsvannnkaellkemqdnvpgktpaqaqlsv 240
QY  241 QTDATATQIEKDGNAIGDAYFAGNAGAVENAKSNNSNISIDSAKAAIAATAKTAQIAEAQ 300
DB  241 qtdatatqiekdgnaigdayfagnagavenaksnnsnidsakaaiaataktqiaeaq 300
QY  301 KKFDPSPILQEAQEMVIOAEKDLKNIPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360
DB  301 kkfpdspilqeaqemviquaekdlknikpadgsdvpnpgtttvggskqggssigsirvsmll 360
QY  361 DDAENETASILMSGFROMIHMFENTENPDSQAQQLAQAQAARAKAGDDSAALADAQK 420
DB  361 ddaenetasilmsgfromihmfentpdsqaqqlaqaqaaraakagddsaaladaqk 420
QY  421 ALEALGKAGQOQGLTALGQIASAAVYVAGVLPQQLVIRARYQVVEQKLISEEDLN 480
DB  421 alealqkagqoqgltnalqgiassavvyvavglpqqvlwiraryqvveqkliseedln 480
QY  481 SAVDHHHHH 490
DB  481 savdhhhhh 490

RESULT 2
AAY71954
ID  AAY71954 standard; Protein; 651 AA.
XX
AC  AAY71954;
XX
DT  26-MAR-2001 (first entry)

```

```

XX  Chlamydia pneumoniae 76 kDa full-length protein.
XX
XX  76 kDa protein; bactericidal; diagnosis; prevention; treatment;
XX  pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
XX  acute respiratory disease; cough; sore throat; hoarseness; fever;
XX  vaccine; immunisation.
XX
XX  Chlamydia pneumoniae.
XX
XX  WO200066739-A2.
XX
XX  09-NOV-2000.
XX
XX  03-MAY-2000; 2000WO-CA00511.
XX
XX  03-MAY-1999; 99US-0132270.
XX
XX  30-JUN-1999; 99US-0141276.
XX
XX  (AVET ) AVENTIS PASTEUR LTD.
XX
XX  Murdin AD, Oomen RP, Wang J, Dunn P;
XX
XX  WPI; 2000-687542/67.
XX
XX  N-PSDB; AAD02063.
XX
XX  Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
XX  useful for vaccinating against Chlamydia infections -
XX
XX  Claim 16a; Fig 1; 112pp; English.
XX
XX  The present sequence is Chlamydia pneumoniae full-length
XX  76 kDa protein. C. pneumoniae 76 kDa protein
XX  is used in the diagnosis, prevention and treatment
XX  of C. pneumoniae infections (e.g. pneumonia, upper respiratory
XX  tract disease, bronchitis, sinusitis and acute respiratory
XX  disease such as cough, sore throat, hoarseness, fever; and
XX  abnormal chest sounds on auscultation). C. pneumoniae sequence
XX  is also used as vaccines for immunising humans against diseases
XX  caused by C. pneumoniae.
XX
XX  Sequence 651 AA;

Query Match          90.7%; Score 2188; DB 21; Length 651;
Best Local Similarity 98.9%; Pred. No. 4.3e-139;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY  1 MNPVIGPGPIDETERTPPADLSAAGLEASAAKSAEQAQRIAGAEAKPKESKTSVERWSI 60
DB  1 mnpvpgpiddetertppadlsaggleasaanksaeeqriagaekpkesktsverwsi 60
QY  61 LRSVAVNLSLADKLGITASSNSSSTSRSDVDSTTATPTPPPTSDYKTAQTAQYDT 120
DB  61 lrsavnlmsladvkltgiasnssstsrsvdvdsttatptppptdyktaqtaqydt 120
QY  121 IFTSTSLADIQAALVSLQDAVTNPKDTAAIDETAIAAEWETKNADAIKVGAQITELAKY 180
DB  121 iftstsladiqaalvslqdvavtnikdtaateetaiaaeawetknadvkvgaqitelaky 180
QY  181 ASDNQAILDLSGLKTSFDLLQTAALQSVANNKKAELLKEMQDNPVPGKTPAQAQSLVD 240
DB  181 asdnqaildlsigkltstfdllqtaalqsvannnkaellkemqdnvpgktpaqaqlsv 240
QY  241 QTDATATQIEKDGNAIGDAYFAGNAGAVENAKSNNSNISIDSAKAAIAATAKTAQIAEAQ 300
DB  241 qtdatatqiekdgnaigdayfagnagavenaksnnsnidsakaaiaataktqiaeaq 300
QY  301 KKFDPSPILQEAQEMVIOAEKDLKNIPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMML 360
DB  301 kkfpdspilqeaqemviquaekdlknikpadgsdvpnpgtttvggskqggssigsirvsmll 360
QY  361 DDAENETASILMSGFROMIHMFENTENPDSQAQQLAQAQAARAKAGDDSAALADAQK 420

```


Db 361 ddaenetasilmgrfmhmtentpdsqaagqelaagaaagaagddsaaladagk 420
 Oy 421 ALEAALGRAGQOQILNALGQIASAAVVSAGVLP 454
 Db 421 aleaalgragqoqilnalgqiasaavvsagvvp 454

RESULT 3
 AAY71956
 ID AAY71956 standard; Protein; 452 AA.
 XX
 AC AAY71956;
 DT 26-MAR-2001 (first entry)
 XX
 DE 3'-truncated Chlamydia pneumoniae 76 kDa protein.
 KW 76 kDa protein; bactericidal; diagnosis; prevention;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;
 KW vaccine; immunisation; treatment; truncation mutant; mutein.
 XX Chlamydia pneumoniae.
 OS Synthetic.
 XX WO200066739-A2.
 PN 09-NOV-2000.
 PD 03-MAY-2000; 2000WO-CA00511.
 PF 03-MAY-1999; 99US-0132270.
 PR 30-JUN-1999; 99US-0141276.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J, Dunn P;
 XX WPI; 2000-687542/67.
 DR N-PSDB; AAD02065.
 XX
 PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 PT useful for vaccinating against Chlamydia infections -
 XX
 PS Claim 16c; Page 104-106; 112pp; English.
 XX
 CC The present sequence is 3'-truncated
 CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
 CC protein is used in the
 CC diagnosis, prevention and treatment of C. pneumoniae
 CC infections (e.g. pneumonia, upper respiratory
 CC tract disease, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX
 SQ Sequence 452 AA;

Query Match 90.5%; Score 2184; DB 21; Length 452;
 Best Local Similarity 99.1%; Pred. No. 5e-139;
 Matches 448; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Oy 1 MWNPIGPGPIDETPTPADLSAQLGSAANKSAEAOIRAGAEAPKESKTDVSERWSI 60
 Db 1 mwnpiGPpIdetertppadlsaqgleaanksaegriaqaeakpkeskdsverwsi 60
 Oy 61 LRSAYNALMSLADKLGIASSNSSSTSRSDVDSTTATPTPPPTSDYKTQATAYDT 120
 Db 61 lrsavnalmsladklgiassnssstsradvdsttatptppptfdyktqataydt 120

Oy 121 IFTSTSLADIAALVSLQDAVTNIKDTAATDEETAIATAEMETKNADAIVKGAQITELAKY 180
 Db 121 iftstsladiqaalvslqdavtnikdtatdeetaiaaewetknadavkvgaqitelaky 180
 Oy 181 ASDNQAILDSLGKLTSLFDLLQTLLOSVAANNKAAELLKEMQDNVVPVKTPAIQSLVD 240
 Db 181 asdnqaildsilgkltslfdllqtllosvannkaellkemqdnvpvpgktpaiqslvd 240
 Oy 241 QTDATATOIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAAKATATATKTOIAEAQ 300
 Db 241 qtdatatatiekdgnaidgdayfagqnasgavenaksnnsidsakaatataktqiaeq 300
 Oy 301 KKPDPSPILQEAEMQVIOAEKDLKNIKPADGSDVPNPQTTVGSKQOQSSIGSIRVSMML 360
 Db 301 kkfdpdpilqeaemviquaedlknikpadgsdvpnpqttvggskqgssigsirvsmll 360
 Oy 361 DDAENETASILMSGFROMIHMENTENPDQAAQOELAAQARAARAKAGDSDSAAALADAOK 420
 Db 361 ddaenetasilmgrfmhmtentpdsqaagqelaagaaagaagddsaaladagk 420
 Oy 421 ALEAALGRAGQOQILNALGQIASAAVVSAGV 452
 Db 421 aleaalgragqoqilnalgqiasaavvsagv 452

RESULT 4
 AAY35358
 ID AAY35358 standard; Protein; 478 AA.
 XX
 AC AAY35358;
 DT 13-SEP-1999 (first entry)
 XX
 DE Chlamydia pneumoniae surface exposed polypeptide.
 KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
 KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
 KW vaccine; neutralising epitope.
 XX Chlamydia pneumoniae.
 XX WO9927105-A2.
 PN 03-JUN-1999.
 PD 20-NOV-1998; 98WO-IB01890.
 PF 04-NOV-1998; 98US-0107078.
 PR 21-NOV-1997; 97FR-0014673.
 XX
 PA (GEST) GENSET.
 XX Griffais R;
 XX WPI; 1999-357842/30.
 XX Genome sequence of Chlamydia pneumoniae
 XX Page 1156; Disclosure; 1912pp; English.
 CC AAY34584-Y35879 represent the proteins encoded by all the open reading
 CC frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
 CC C. pneumoniae causes respiratory disease such as pneumonia and
 CC bronchitis and is thought to be a contributing factor in heart
 CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
 CC nodosum or pharyngitis. The polypeptides encoded by the open reading
 CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
 CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
 CC nucleotide sequences can also be used as immunogenic compositions,
 CC especially where the vector directs the expression of a neutralising
 CC epitope of C. pneumoniae.
 XX Sequence 478 AA;

Query Match 89.6%; Score 2162; DB 20; Length 478;
 Best Local Similarity 99.1%; Pred. No. 1.6e-137;
 Matches 443; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MYNPIGPGIDFTERTPPADLSAQGLEASAAKSAEAGRIAGAEAKPKESKTDSDVERWSI 60
 DB 3 mvnplgpgpideterppadlsaqgleasaanksaegriagaeakpkeskdsvrws 62
 QY 61 LRSAYNALMSADKLGIASSSSSTSRSDVSTTATAPPPPTSDDYKTQATAYDT 120
 DB 63 lrsaynalmsadkligiassssstsrsvdsttatppppptfddyktaqaydt 122
 QY 121 IFTSLADIQAALVSLQDAVNIKDTAATDETAIAAEWETKNADAIVKGAQITELAKY 180
 DB 123 iftstladiqaalvslqdvavtnikdtatdeetaiaaewetknadavkvgaqitelaky 182
 QY 181 ASDNQAILDSGLKLSFDLLQTLQSVANNKAAELKEMQDNVPVFGKTPAIAQSLVD 240
 DB 183 asdnqaildsiglksfdllqallqsvannnkaaellkemqdnvpvpgktpaiagslvd 242
 QY 241 OTDATATQIEKDGNAIGAYFAGONASGAVENAKSNNSISNIDSKAAIAIATAKTOIAEAQ 300
 DB 243 qdatatqiekdgnaigayfagongasgavenaksnnsidskaaiataktqiaeaq 302
 QY 301 KFPDPSPIQEAQMVIAQEKDLNKPADGSDVPNPGTTVGSKQOGSSIGSIRVSM 360
 DB 303 kfpdspilqeaqmviaekdlknkpadgsdvpnppttvggskqogssigsirvsm 362
 QY 361 DDAENETASILMSGFRQIMHFNTEPNDSQAQOELAAQARAAGDSDSAAALADAK 420
 DB 363 ddaenetasilmsgfrqimhfnntepdsqaqgelaqaraagaagdsdaaaladagk 422
 QY 421 ALEAALGRAGQOQILNALGOIASAAV 447
 DB 423 aleaalgragqqqilnalgiasaav 449

RESULT 5
 AAY71955
 ID AAY71955 standard; Protein; 583 AA.
 XX
 AC AAY71955;
 DT 26-MAR-2001 (first entry)
 XX
 DE 5'-truncated Chlamydia pneumoniae 76 kDa protein.
 XX
 KW 76 kDa protein; bactericidal; diagnosis; prevention;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;
 KW vaccine; immunisation; treatment; truncation mutant; mutein.
 XX
 OS Chlamydia pneumoniae.
 OS Synthetic.
 XX
 FH Key Location/Qualifiers
 FT Misc-difference 497
 FT /note= "Encoded by ACT"
 XX
 FN WO200066739-A2.
 XX
 PD 09-NOV-2000.
 XX
 PE 03-MAY-2000; 2000WO-CA00511.
 XX
 PR 03-MAY-1999; 99US-0132270.
 PR 30-JUN-1999; 99US-0141276.
 XX
 PA (AVET) AVENTIS PASTEUR LTD.
 XX
 PI Murdin AD, Oomen RP, Wang J, Dunn P;

XX WPI: 2000-687542/67.
 DR N-PSDB; AAD02064.
 XX
 PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 useful for vaccinating against Chlamydia infections -
 XX
 PS Claim 16b; Page 100-102; 112pp; English.
 XX
 CC The present sequence is 5'-truncated
 CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
 CC protein is used in the
 CC diagnosis, prevention and treatment of C. pneumoniae
 CC infections (e.g. pneumonia, upper respiratory
 CC tract disease, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX
 SQ Sequence 583 AA;

Query Match 76.7%; Score 1850; DB 21; Length 583;
 Best Local Similarity 98.7%; Pred. No. 2e-116;
 Matches 381; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 69 MSLADKLGIASSSSSTSRSDVSTTATAPPPPTSDDYKTQATAYDTFTSTSLA 128
 DB 1 msladkligiassssstsrsvdsttatppppptfddyktaqaydtftstsla 60
 QY 129 DIOAALVSLQDAVNIKDTAATDETAIAAEWETKNADAIVKGAQITELAKYASDNOATL 188
 DB 61 diqaalvslqdvavtnikdtatdeetaiaaewetknadavkvgaqitelakysdnqall 120
 QY 189 DSLGKLTSPDLLQTLQSVANNKAAELKEMQDNVPVPGKTPAIAQSLVDQTDATATQ 248
 DB 121 dslgkltspdllqallqsvannnkaaellkemqdnvpvpgktpaiagslvdqtdatq 180
 QY 249 IEKDGNAIGAYFAGONASGAVENAKSNNSISNIDSKAAIAIATAKTOIAEAOKFPDPSPI 308
 DB 181 iekdgnaigayfagongasgavenaksnnsidskaaiataktqiaeaokfpdsp 240
 QY 309 LQEAQMVIAQEKDLNKPADGSDVPNPGTTVGSKQOGSSIGSIRVSMILDDAENETA 368
 DB 241 lqeaqmviaekdlknkpadgsdvpnppttvggskqogssigsirvsmllddaeneta 300
 QY 369 SILMSGFRQIMHFNTEPNDSQAQOELAAQARAAGDSDSAAALADAKALEAALGK 428
 DB 301 silmsgfrqimhfnntepdsqaqgelaqaraagaagdsdaaaladakaaleaalgk 360
 QY 429 AGOQOQILNALGOIASAAVVSAGVLP 454
 DB 361 agqqqgilnalgiasaavvsagvlp 386

RESULT 6
 AAY37571
 ID AAY37571 standard; Protein; 350 AA.
 XX
 AC AAY37571;
 DT 07-OCT-1999 (first entry)
 XX
 DE Chlamydia trachomatis surface exposed protein.
 XX
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihhepatitis;
 KW nongonococcal urethritis; epidymitis; cervicitis; salpingitis;
 KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.
 XX
 OS Chlamydia trachomatis.
 OS

PN W09928475-A2.
 XX 10-JUN-1999.
 PD
 XX 27-NOV-1998; 98WO-IB01939.
 XX
 PF 04-NOV-1998; 98US-0107077.
 PR 28-NOV-1997; 97FR-0015041.
 PR 17-DEC-1997; 97FR-0016034.
 XX
 PA (GEST) GENSET.
 XX
 XX Griffais R;
 PI
 XX WPI; 1999-371125/31.
 DR
 XX
 XX Genome sequence of Chlamydia trachomatis
 PT
 XX Disclosure; Page 1226-1227; 1755pp; English.
 XX
 XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
 CC of Chlamydia trachomatis (see AZ01425). The polypeptides can be used as
 CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
 CC can also be used to control growth of the microorganism. Chlamydia
 CC trachomatis is responsible for a large number of diseases, e.g. eye
 CC diseases such as conventional trachoma, nonendemic trachoma,
 CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
 CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
 CC perihepatitis, Bartholinitis; pneumopathy in breast feeding infants;
 CC and venereal lymphogranulomatosis. The polypeptides of the invention
 CC may be of use in treating these diseases.
 XX
 XX Sequence 350 AA;
 SQ
 Query Match 11.7%; Score 281.5; DB 20; Length 350;
 Best Local Similarity 24.9%; Pred. No. 2.6e-11;
 Matches 95; Conservative 71; Mismatches 122; Indels 93; Gaps 13;
 QY 2 VNPIGPGPIDETERTPPADLSAQGLEASAAKSAEAOIAEAKPKESKTDTSVERWSIL 61
 Db 23 mnpingqi-----asnetkestasea-----ssassssvswsf 60
 QY 62 RSAYNALMSLADKLGIASSNSSTSRSDVDSTATPTPPPTSDDYKTAQTAYDTI 121
 Db 61 ssaknalislrd--ailnknssptdsis-qleatsts-tvtrvaakdy-dkaksnfda 115
 QY 122 FT-----STSLADIQALVSLQ-----DAVTNKDTAATDEETAIAAEW 160
 Db 116 ksglenaktlaeyletkmadlmaalqdmansdpsndhteelnikkaleaqkdti----- 170
 QY 161 ETKNADAIKVGAIQITELAKYASDNOAILDSLGKLTSPFLLQTLLOSVAANNKAAELIKE 220
 Db 171 -----dkinklvtnqknslteakltkdsadqipainrleinknsahqilke 219
 QY 221 MODNPVPEGKTPAIQAISLVDTATATQIEKD-----GNAIGDAYFAGONASGA 269
 Db 220 lkeq-----isnykavltdevkivefseagiklgalqsvdgdqgsaa 265
 QY 270 VENAKSNNSISNIDSAKAAITAKTQIAEAO---KKFPDPSILQAEOMVTOAEKDLKNI 326
 Db 266 vlqarqnsnpdniaatklidaaktvnelkqehqeadspdvkkaeqinqagqdiqti 325
 QY 327 KPAGSDVDPNPGTTVGGSKQQ 347
 Db 326 tps-gldipivgpgsgxpqe 345
 RESULT 7
 ID AAB13695
 XX AAB13695 standard; Protein; 361 AA.
 AC AAB13695;

XX 02-FEB-2001 (first entry)
 DT
 XX Chlamydia sp. protein # 6.
 DE
 XX Chlamydial infection; sexually transmitted disease;
 KW pelvic inflammatory disease; PID; tubal obstruction; infertility;
 KW trachoma; blindness; acute respiratory tract infection;
 KW atherosclerosis; coronary heart disease; antibacterial.
 XX
 OS Chlamydia sp..
 XX
 XX WO200034483-A2.
 PN
 XX 15-JUN-2000.
 PD
 XX 08-DEC-1999; 99WO-US29012.
 PF
 XX 08-DEC-1999; 98US-0208277.
 PR 08-APR-1999; 99US-0288594.
 PR 01-OCT-1999; 99US-0410568.
 PR 22-OCT-1999; 99US-0426571.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 XX Probst P, Bhatia A, Skeiky YAW, Fling SP, Jen S, Stromberg EJ;
 PI
 XX WPI; 2000-431303/37.
 DR
 XX Isolated polypeptide for diagnosis and treatment of Chlamydia infection
 XX comprises immunogenic portion of Chlamydia antigen, which comprises
 PT amino acid sequence encoded by polynucleotide sequence -
 PT
 XX Disclosure; Pages 251-253; 256pp; English.
 PS
 XX The present invention relates to new nucleic acid sequences and the
 CC proteins encoded by the nucleic acid sequences. The encoded proteins
 CC comprise an immunogenic portion of a Chlamydia antigen. The encoded
 CC proteins are useful for the serodiagnosis and treatment of Chlamydia
 CC infection. Chlamydiae are intracellular bacterial pathogens that are
 CC responsible for a wide variety of human infections. C. trachomatis
 CC infection is one of the most common sexually transmitted diseases and can
 CC lead to pelvic inflammatory disease (PID), resulting in tubal obstruction
 CC and infertility. Trachoma due to ocular infection with C. trachomatis is
 CC the leading cause of preventable blindness worldwide. C. pneumoniae is a
 CC major cause of acute respiratory tract infections in humans and is also
 CC thought to play a role in the pathogenesis of atherosclerosis and
 CC coronary heart disease. The present sequence is a protein isolated in the
 CC present invention.
 CC Sequence 361 AA;
 SQ
 Query Match 9.6%; Score 231.5; DB 21; Length 361;
 Best Local Similarity 35.4%; Pred. No. 6.3e-08;
 Matches 57; Conservative 30; Mismatches 61; Indels 13; Gaps 5;
 QY 301 KKFPDPSILQAEOMVIOAEKDLKNIKPADGSDVDPNPGTTVGGSKQSSSTGSI----- 354
 Db 2 qeiadspvlkkaeqinqagqdiqtips-gldipivgpgs-gsaasagaaalkssnns 59
 QY 355 -RVSMMLDDAENETASILMSGFROMIHMFNTENPDSSAAQOELAAQARAAK---AAGDSD 410
 Db 60 grisllldvdmamaaiaqgmfrsmieqfvrnnpatakelameqltamsdqlvgadqe 119
 QY 411 AAAALADAAQKALEALGKAGQQGILNALGQITASAAVVSAG 451
 Db 120 lpaelqaiakdalaqal-kpsadglatamgvaataakvvg 159
 RESULT 8
 ID AAG83263
 XX AAG83263 standard; Protein; 361 AA.

Sequence 331 AA;

31

RESULT 17

AC AAY81609;

DT 24-MAY-2000 (first entry)

DE
Stren

Streptococcus pneumoniae type 4 protein sequence #109.

AA Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.

OS Streptococcus pneumoniae.

XX PN WO200005737-A2

10-FEB-2000 00:XX

XX
27 - TIT - 1000. 00000-00000000

1
 2
 3
 4
 5
 6
 7
 8
 9
 10
 11
 12
 13
 14
 15
 16
 17
 18
 19
 20
 21
 22
 23
 24
 25
 26
 27
 28
 29
 30
 31
 32
 33
 34
 35
 36
 37
 38
 39
 40
 41
 42
 43
 44
 45
 46
 47
 48
 49
 50
 51
 52
 53
 54
 55
 56
 57
 58
 59
 60
 61
 62
 63
 64
 65
 66
 67
 68
 69
 70
 71
 72
 73
 74
 75
 76
 77
 78
 79
 80
 81
 82
 83
 84
 85
 86
 87
 88
 89
 90
 91
 92
 93
 94
 95
 96
 97
 98
 99
 100
 101
 102
 103
 104
 105
 106
 107
 108
 109
 110
 111
 112
 113
 114
 115
 116
 117
 118
 119
 120
 121
 122
 123
 124
 125
 126
 127
 128
 129
 130
 131
 132
 133
 134
 135
 136
 137
 138
 139
 140
 141
 142
 143
 144
 145
 146
 147
 148
 149
 150
 151
 152
 153
 154
 155
 156
 157
 158
 159
 160
 161
 162
 163
 164
 165
 166
 167
 168
 169
 170
 171
 172
 173
 174
 175
 176
 177
 178
 179
 180
 181
 182
 183
 184
 185
 186
 187
 188
 189
 190
 191
 192
 193
 194
 195
 196
 197
 198
 199
 200
 201
 202
 203
 204
 205
 206
 207
 208
 209
 210
 211
 212
 213
 214
 215
 216
 217
 218
 219
 220
 221
 222
 223
 224
 225
 226
 227
 228
 229
 230
 231
 232
 233
 234
 235
 236
 237
 238
 239
 240
 241
 242
 243
 244
 245
 246
 247
 248
 249
 250
 251
 252
 253
 254
 255
 256
 257
 258
 259
 260
 261
 262
 263
 264
 265
 266
 267
 268
 269
 270
 271
 272
 273
 274
 275
 276
 277
 278
 279
 280
 281
 282
 283
 284
 285
 286
 287
 288
 289
 290
 291
 292
 293
 294
 295
 296
 297
 298
 299
 300
 301
 302
 303
 304
 305
 306
 307
 308
 309
 310
 311
 312
 313
 314
 315
 316
 317
 318
 319
 320
 321
 322
 323
 324
 325
 326
 327
 328
 329
 330
 331
 332
 333
 334
 335
 336
 337
 338
 339
 340
 341
 342
 343
 344
 345
 346
 347
 348
 349
 350
 351
 352
 353
 354
 355
 356
 357
 358
 359
 360
 361
 362
 363
 364
 365
 366
 367
 368
 369
 370
 371
 372
 373
 374
 375
 376
 377
 378
 379
 380
 381
 382
 383
 384
 385
 386
 387
 388
 389
 390
 391
 392
 393
 394
 395
 396
 397
 398
 399
 400
 401
 402
 403
 404
 405
 406
 407
 408
 409
 410
 411
 412
 413
 414
 415
 416
 417
 418
 419
 420
 421
 422
 423
 424
 425
 426
 427
 428
 429
 430
 431
 432
 433
 434
 435
 436
 437
 438
 439
 440
 441
 442
 443
 444
 445
 446
 447
 448
 449
 450
 451
 452
 453
 454
 455
 456
 457
 458
 459
 460
 461
 462
 463
 464
 465
 466
 467
 468
 469
 470
 471
 472
 473
 474
 475
 476
 477
 478
 479
 480
 481
 482
 483
 484
 485
 486
 487
 488
 489
 490
 491
 492
 493
 494
 495
 496
 497
 498
 499
 500
 501
 502
 503
 504
 505
 506
 507
 508
 509
 510
 511
 512
 513
 514
 515
 516
 517
 518
 519
 520
 521
 522
 523
 524
 525

PR 19-MAR-1999; 99US-0125164.

PA (MICR-) MICROBIAL TECHNICS LTD.

PI Gilbert CFG, Hansbro PM;

DR WPI; 2000-195300/17.

New Streptococcal protein, useful as a vaccine, for diagnosis of pneumococcal diseases and for screening agents capable of antagonizing or inhibiting expression of the protein -

PS Claim 1: Page 83: 108pp: English.

AA081501 to AA081679 represent specifically claimed protein sequences isolated from *Streptococcus pneumoniae*. AA0815407 to AA0815590 represent specifically claimed nucleotide sequences isolated from *S. pneumoniae*. The sequences have antibacterial and antiinflammatory properties. The protein sequences, and fragments of them, are useful as immunogens and/or antigens. The nucleotide sequences can be used in vaccines and diagnostic assays. The proteins and nucleotides can be useful for the detection and diagnosis of *S. pneumoniae*. The protein sequences are also useful for screening an agent capable of antagonising, inhibiting or

RESULT 10

AAY37572
ID AAY37572 standard; Protein; 331 AA.

AC AAY37572:

AA	DT	07-OCT-1999	(first entry)

Chlamydia trachomatis cellular envelope protein.

AA	Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW	paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW	nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW	bartholinitis; pneumopathy; venereal lymphogranulomatosis.

OS Chlamydia trachomatis.

XX PN WO9928475-A2.

XX
PD
10-JUN-1999

XX
27-NOV-1998. 09WQ-TP01030

XX
04 - NOV - 1998. 0905 - 0107077

PR 28-NOV-1997; 97ER-0015041.

XX

XX
XXXX
XX
NY CTPTTTT 11

DR WFL; 1999-3/1125/31.
XX

Genome sequence of *Chlamydia trachomatis* P1

PS Disclosure; Page 1227-1228; 1755pp; English.
vv

CC AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome

XX DE S. mutans antigen I/II.
 XX KW Caries; antigen I/II; epitope; tooth decay; vaccine; immunisation.
 XX OS Streptococcus mutans.
 XX FH Key Location/Qualifiers
 XX FT Misc-difference 618..650
 XX FT /note= "amino acids 618-650 differ from the
 XX FT residues deduced from the nucleic acid
 XX FT sequence."
 XX PN WO9623886-A1.
 XX PD 08-AUG-1996.
 XX PF 31-JAN-1996; 96WO-GB00207.
 XX PR 31-JAN-1995; 95GB-0001826.
 XX PA (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 XX PI Kelly C, Lehner T;
 XX DR WPI; 1996-371434/37.
 XX DR N-PSDB; AAT36122.
 XX KW Nucleic acid encoding polypeptide for prevention or treatment of
 XX PT dental caries - which stimulates T or B cell response, and/or
 XX PT adheres to tooth in competition with Streptococcus mutans antigen
 XX PT I/II
 XX PS Disclosure; Page 45-46; 63pp; English.
 XX CC Streptococcus mutans antigen I/II (AAW02098) is a 185 kDa cell surface
 XX CC protein at least partly responsible for S. mutans adhesion to teeth.
 XX CC The I/II antigen includes a series of overlapping T-cell, B-cell and
 XX CC adhesion epitopes. Fragments (see also AAT36111-21) of the I/II
 XX CC antigen gene (see also AAT36122) can be used to produce recombinant
 XX CC polypeptides (AAW02087-97) carrying such epitopes for use in vaccines
 XX CC for immunisation against dental caries.
 XX SO Sequence 1561 AA;
 Query Match 6.3%; Score 152; DB 17; Length 1561;
 Best Local Similarity 22.3%; Pred. No. 0.093;
 Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;
 Qy 48 KESKTSVERWSILRSVAVNALMSLADKLGIASSN-SSSSTSRSDVDSSTTATPTPPPT 106
 Db 2 kvktygrkskiskitcigavltgvaavsvagkvfadetttsdvdkvvgvtgtnpat 61
 Qy 107 -----SDYKTKQAQYATYTIPTST-----SLADIQAALVS-LQDVTNPKDT 147
 Db 62 nlpeaqsasqaeqsgtklerqmvhtievkptldlqaaakdaksagvnnvqgdvkn-kgt 120
 Qy 148 AATDEETAIAREWE-----TKNADAIK-----VCAQITELAKYASDQAIILSLGK 193
 Db 121 vkttee-avqketeikedytkqaeidkkttdqykdsvaaheaevakikaknqatkeqygk 179
 Qy 194 LTFSDLL-QTALLOSVAANNKAAEL-----LKEMQDNVPVPGKYPATIAQSLVDQTDATAT 247
 Db 180 ----dmvahaeverinaanaasktayeklaqyqadlaavqktnaanqasy-qkalaay 234
 Qy 248 QIE-----KGNAGIDAVFAGNAGAVENAKSNNSISNIDSAKAIATATK-----TQIAEA 299
 Db 235 qaelkrvqeanaaakaaqy---dtavaannakntelaanaeairkrnakataeyetklaqy 291
 Qy 300 Q---KKFPDPSILOFAE-QWVIOA-EKDLKNIKPAD-----GSDVPNPGTIVG 342
 Db 292 qaelkrvqeananeadyqakityqtelarvqkanadakaayaevaannaknaaltcae 351

Qy 343 GS-----KQGGSSIGSIRVSMLLDDAENETASTILMSGFRQIMHFNTE 385
 Db 352 ntaikqrnenakatyaaalkyeadlaavkkanaaneadyqak---ltaqtelarvqka 408
 Qy 386 NPDQAQOQELAAQARAKAA--GDSSAAAAL-ADAQKALEAALGK 428
 Db 409 nedakaayeaavaanaanaaltaentaikkrnadakadyeaklak 454
 RESULT 14
 AAR84568
 ID AAR84568 standard; Protein; 643 AA.
 XX AC AAR84568;
 XX DT 09-MAY-1996 (first entry)
 XX DE Trypanosoma cruzi TCR27 polypeptide, Ag15.
 XX KW TCR27; Chagas disease; repeat unit; diagnosis; blood screening;
 XX KW recombinant; fusion protein; glutathione-S-transferase.
 XX OS Trypanosoma cruzi.
 XX FH Key Location/Qualifiers
 XX FT Region 329..552
 XX FT /label= repeat_region
 XX FT /note= "16 of 69 repeat units of 14 amino acids"
 XX PN WO9525797-A1.
 XX PD 28-SEP-1995.
 XX PF 20-MAR-1995; 95WO-US03191.
 XX PR 24-MAR-1994; 94US-0216894.
 XX PA (KIRC/) KIRCHHOFF L V.
 XX PA (OTSU/) OTSU K.
 XX PI Kirchoff LV, Otsu K;
 XX DR WPI; 1995-344618/44.
 XX DR N-PSDB; AAT05332.
 XX PT New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as
 XX PT immunoassay reagent for specific diagnosis of Chagas disease, also
 XX PT related nucleic acid and transformed cells
 XX PS Disclosure; Page 40-41; 68pp; English.
 XX CC AAR84565-R84569 are polypeptides of the TCR27 protein of T.cruzi
 XX CC The proteins are all fusion products with glutathione-S-transferase
 XX CC (GST) and some contain a linker sequence. The TCR27 protein comprises
 XX CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly
 XX CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence
 XX CC encodes the GST sequence, the Ag44 polypeptide contg. 16 of the 69
 XX CC repeat units and also contains the amino and carboxy terminal
 XX CC peptides of TCR27. The TCR27 polypeptides of the invention are useful
 XX CC for the diagnosis of Chagas disease (American Trypanosomiasis), they
 XX CC are capable of detecting anti-T.cruzi antibodies; or for blood
 XX CC screening. The TCR27 protein has epitopes to which most T.cruzi
 XX CC infected individuals have antibodies. The TCR27 polypeptides will not
 XX CC react with serum from patients with leishmaniasis, schistosomiasis,
 XX CC or autoimmune disease and are hence less likely to cause false
 XX CC positives in diagnosis.
 XX SO Sequence 643 AA;

Query Match 6.2%; Score 150.5; DB 16; Length 643;
 Best Local Similarity 21.5%; Pred. No. 0.037;

PR	30-APR-1999;	99US-01332407;
PR	04-MAY-1999;	99US-01332484;
PR	05-MAY-1999;	99US-01332485;
PR	06-MAY-1999;	99US-01332486;
PR	06-MAY-1999;	99US-01332487;
PR	07-MAY-1999;	99US-01332863;
PR	11-MAY-1999;	99US-01334256;
PR	17-MAY-1999;	99US-01342118;
PR	14-MAY-1999;	99US-01342119;
PR	14-MAY-1999;	99US-01342211;
PR	14-MAY-1999;	99US-01343701;
PR	18-MAY-1999;	99US-01343768;
PR	19-MAY-1999;	99US-01349491;
PR	20-MAY-1999;	99US-01351124;
PR	21-MAY-1999;	99US-01353533;
PR	24-MAY-1999;	99US-01356629;
PR	25-MAY-1999;	99US-01360211;
PR	27-MAY-1999;	99US-01363632;
PR	28-MAY-1999;	99US-01367682;
PR	01-JUN-1999;	99US-01372222;
PR	03-JUN-1999;	99US-01375528;
PR	04-JUN-1999;	99US-01375002;
PR	07-JUN-1999;	99US-01377724;
PR	08-JUN-1999;	99US-01380094;
PR	10-JUN-1999;	99US-01385450;
PR	10-JUN-1999;	99US-01388847;
PR	14-JUN-1999;	99US-01391119;
PR	16-JUN-1999;	99US-01394552;
PR	16-JUN-1999;	99US-01394553;
PR	17-JUN-1999;	99US-01394942;
PR	18-JUN-1999;	99US-01394954;
PR	18-JUN-1999;	99US-01394632;
PR	18-JUN-1999;	99US-01394633;
PR	18-JUN-1999;	99US-01394556;
PR	18-JUN-1999;	99US-01394578;
PR	18-JUN-1999;	99US-01394558;
PR	18-JUN-1999;	99US-01394559;
PR	18-JUN-1999;	99US-01394600;
PR	18-JUN-1999;	99US-01394611;
PR	18-JUN-1999;	99US-01394632;
PR	18-JUN-1999;	99US-01394633;
PR	18-JUN-1999;	99US-01397550;
PR	18-JUN-1999;	99US-01397563;
PR	21-JUN-1999;	99US-01398137;
PR	22-JUN-1999;	99US-01398699;
PR	23-JUN-1999;	99US-01403553;
PR	23-JUN-1999;	99US-01403554;
PR	24-JUN-1999;	99US-01406955;
PR	28-JUN-1999;	99US-01408233;
PR	28-JUN-1999;	99US-01409911;
PR	29-JUN-1999;	99US-01412827;
PR	30-JUN-1999;	99US-01418422;
PR	01-JUL-1999;	99US-01429200;
PR	01-JUL-1999;	99US-01429777;
PR	02-JUL-1999;	99US-01421594;
PR	02-JUL-1999;	99US-01420555;
PR	06-JUL-1999;	99US-01423900;
PR	08-JUL-1999;	99US-01428033;
PR	09-JUL-1999;	99US-01429200;
PR	12-JUL-1999;	99US-01435422;
PR	13-JUL-1999;	99US-01435424;
PR	14-JUL-1999;	99US-01436244;
PR	15-JUL-1999;	99US-01440055;
PR	16-JUL-1999;	99US-01440085;
PR	16-JUL-1999;	99US-01440826;
PR	19-JUL-1999;	99US-01443311;
PR	19-JUL-1999;	99US-01443331;
PR	19-JUL-1999;	99US-01443332;
PR	19-JUL-1999;	99US-01443333;
PR	19-JUL-1999;	99US-01443352;
PR	20-JUL-1999;	99US-01446322;
PR	20-JUL-1999;	99US-01448884;
PR	21-JUL-1999;	99US-01448414;
PR	21-JUL-1999;	99US-01450866;


```
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145931.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147302.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147433.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149388.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161922.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.
Query Match 6.1%; Score 148; DB 21; Length 571;
Best Local Similarity 20.8%; Pred. No. 0.047;
Matches 94; Conservative 82; Mismatches 167; Indels 108; Gaps 20;
Qy 77 IASNSSTSRSDVSTTATATPPPTSDDYKTOATYDTFTSTSLADIOAALVS 136
Db 87 lgasgskasts-----sssssetlplvtrdhntra-----lstdlvsnmmltt 132
Qy 137 LQ---DAVTNIKDTAATDEETAIAEWE-----TKNADAIKVCAQITE--- 176
Db 133 lqpnvdpvchkhleeseieigsvkkargkengnydlfgrcfdvtsna---kigldvpeddi 189
Qy 177 ---LAKYASDNQAIILDSLGKLTSLFLLQTLALQSVANNKKAELKEMQDNPPVPGKTPA 233
Db 190 asvvsflsldnvnklasnfwedlglfditw-----nqaenavskivds--tpgdmt 237
Qy 234 IAQ-----SLVDDQTDATATQTEKDGNAIGDAYFRAGQNASGA-----VENAKS 275
Db 238 ttqggscfcvghgstnlvdsnrsfsgngdteigdafstshvcsvdqilediedaks 297
Qy 276 N--NSISNIDSAKAIAATAKTQIAFAOKKFPDS-----PILQEAQOMVIOAEKDLKNTKP 328
Db 298 nkhlitemetvtnimreveikekdaekseearggldtqkveel-----kkmlelake 353
Qy 329 ADGSDVPNPGTGVGSKQOGSSIGSIRVSMI--LDDAENETASIL--MSGFRQMIHMFNTE 385
Db 354 andm---hagevygeksilatevkelennllnlseerkslaildemrgslei-----403
Qy 386 NPDQAAQOELAAQARAARAKAGDDSAALADAKALEAAL--GKAGOOQGLNA-----438
Db 404 ---rlaaalelkktaekkkedsalkalaeqeanmekvqvgeskllqgeaeenskldrf 460
Qy 439 ---LQOIASAAVWSAGVLPLOQVLWIRARYO 466
Db 461 lmdrgqlvdtlqgeisvi-cqdvkillekfe 490
Search completed: February 7, 2002, 21:34:36
Job time: 20755 sec
```


GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: February 7, 2002, 21:36:13 ; Search time 85.91 seconds
(without alignments)
128.351 Million cell updates/sec

Title: US-09-391-606-8
Perfect score: 2412
Sequence: 1 MNPICGPIDETRTPAD.....QKLISDELNSAVDHHHHH 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCTUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	152	6.3	1561	3	US-08-894-017-23
2	150.5	6.2	643	2	US-08-216-894-8
3	150.5	6.2	643	4	US-09-115-746-8
4	147	6.1	1566	2	US-08-687-956A-23
5	146.5	6.1	1565	6	5352450-2
6	138.5	5.7	564	2	US-08-216-894-2
7	138.5	5.7	564	4	US-09-115-746-2
8	138	5.7	2842	1	US-07-741-940-7
9	138	5.7	2842	1	US-08-289-548A-7
10	138	5.7	2842	1	US-08-452-654-7
11	138	5.7	2843	1	US-07-741-940-2
12	138	5.7	2843	1	US-08-289-548A-2
13	138	5.7	2843	1	US-08-452-654-2
14	138	5.7	2843	1	US-08-452-655B-2
15	138	5.7	2843	1	US-08-452-655B-7
16	138	5.7	2843	2	US-08-370-235A-2
17	138	5.7	2843	3	US-08-450-582-2
18	138	5.7	2843	3	US-08-450-582-7
19	138	5.7	2973	2	US-08-821-355A-7
20	138	5.7	2973	2	US-09-003-687A-7
21	138	5.7	2973	4	US-09-136-605-7
22	136.5	5.7	1154	4	US-08-974-549A-611
23	136.5	5.7	2285	4	US-09-308-375-2
24	135.5	5.6	639	1	US-08-466-390-2
25	135.5	5.6	639	1	US-08-470-950-2
26	135.5	5.6	639	1	US-08-467-781-2
27	135.5	5.6	639	1	US-08-195-487-2

28	135.5	5.6	639	2	US-08-483-924-2	Sequence 2, Appli
29	135.5	5.6	639	5	PCT-US93-06160-2	Sequence 2, Appli
30	135	5.6	885	2	US-08-533-306A-4	Sequence 4, Appli
31	135	5.6	885	2	US-08-742-923A-4	Sequence 4, Appli
32	131.5	5.5	593	2	US-08-591-079-8	Sequence 8, Appli
33	131.5	5.5	593	2	US-08-591-079-10	Sequence 10, Appli
34	130	5.4	2035	1	US-08-046-585-5	Sequence 5, Appli
35	130	5.4	2035	1	US-08-393-703-5	Sequence 5, Appli
36	130	5.4	2035	5	PCT-US93-11721-5	Sequence 5, Appli
37	129.5	5.4	816	2	US-08-533-306A-6	Sequence 6, Appli
38	129.5	5.4	816	2	US-08-742-923A-6	Sequence 6, Appli
39	127.5	5.3	635	4	US-08-931-608A-5	Sequence 5, Appli
40	126.5	5.2	1093	5	PCT-US93-03077-1	Sequence 1, Appli
41	126	5.2	466	4	US-08-235-836C-107	Sequence 107, App
42	125	5.2	2101	1	US-08-466-390-4	Sequence 4, Appli
43	125	5.2	2101	1	US-08-470-950-4	Sequence 4, Appli
44	125	5.2	2101	1	US-08-467-781-4	Sequence 4, Appli
45	125	5.2	2101	1	US-08-195-487-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1
US-08-894-017-23
; Sequence 23, Application US/08894017
; Patent No. 6024958
; GENERAL INFORMATION:
; APPLICANT: Lehner, Thomas
; TITLE OF INVENTION: POLYPEPTIDE FRAGMENTS CAPABLE
; OF COMPETITION WITH STREPTOCOCCUS MUTANS ANTIGEN I/II
; NUMBER OF SEQUENCES: 27
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: MORRISON & FOERSTER
; STREET: 2000 Pennsylvania Avenue, NW, suite 5500
; CITY: Washington
; STATE: DC
; COUNTRY: USA
; ZIP: 20006-1888
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows
; SOFTWARE: FASTSEQ for Windows Version 2.0b
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/894,017
; FILING DATE: 31-JUL-1997
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/GB96/00207
; FILING DATE: 31-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Murashige, Kate H
; REGISTRATION NUMBER: 29,959
; REFERENCE/DOCKET NUMBER: 25150-20067.00
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-887-1500
; TELEFAX: 202-822-0168
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1561 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-894-017-23

Query Match 6.3%; Score 152; DB 3; Length 1561;
Best Local Similarity 22.3%; Pred. No. 0.004;
Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;

```

QY 48 KESTDTSVERWSILRSVAVNALMSLADKLGITASSN-SSSTSRSDVDSTTATPTPPPT 106
Db 2 KVKTYGFRKSKISKTLCGVLGTVAAVSVAGQKVFADETTTSDVTKVVGQTGNPAT 61
QY 107 -----SDYKTKQATAYDTIFTST-----SLADIAALYS-LQDAVNIKDT 147
Db 62 NLPEAGSASQASQSQKLERQWVHTIEVPTDQAAKDAKSAGVNVQDADVN-KGT 120
QY 148 AATDEETAIAAEWE-----TKNADAIAK-----YCAQITELAKYASNDQAILSLGK 193
Db 121 VKTAE-AYKETEIKEDYTKQAEIKTKTDQYKSDVAAHEAEVAKIKAKNOATKEGYK 179
QY 194 LTFSDLL-OTALLQSVANNKAAEL-----LKEMQDNPVPGKTPATAQSLVDQTDATAT 247
Db 180 ----DMVAHKAEVERINAANAASYAYEAKLAQYADLAQVAAVOKTNAANQASY-OKALAA 234
QY 248 QIE-----KDGNAIGDAYFAGNAGAVENAKSNNSISNIDSAKAAIATAK-----TQIAEA 299
Db 235 QAEKRVQEAANAAKAAV---DTAVANNAKNTETAANEELIRKRNATAKAEYETKLAQY 291
QY 300 Q-----KFFPDSPILOQAE-OMVIOA-EKDLNKNIPAD-----GSDVNPNGTGVG 342
Db 292 QAEKRVQEAANAADYQAKLTAYQTELARVOKANADAKAAEYAAVAAANNAKNAALTAE 351
QY 343 GS-----KQGSIGSIRVSMILLDDAENETASILMSGFRQMIHMFNTE 385
Db 352 NTAIKORNENAKATYEAALQYADLAQVAAVOKNANAENADYQAK---LTAYTELARVOKA 408
QY 386 NPDSQAQOELAAQARAATAKAA--GDSSAAAAL-ADAQKALEAALGK 428
Db 409 NADAKAAYEAANAANAALTAENTAIKKRNADAKADYEAALAK 454

```

RESULT 2

```

US-08-216-894-8
; Sequence 8, Application US/08216894
; Patent No. 5876734
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/216,894
; FILING DATE: 24-MAR-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 643 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-216-894-8

```

```

Query Match 6.2%; Score 150.5; DB 2; Length 643;
Best Local Similarity 21.5%; Pred. No. 0.0015;
Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;
QY 7 PGIDETERTPPADLSAAGLEASAAKSAQAQIAGAAKPKESKTSVERWSILRSVAVN 66
Db 216 PPKSDDLPGSPSQL-QQAENNTSKKEMTKLRKVKAKEKLDINRAKLEERN 273
QY 67 ALMSLADKLGITASSNSSSTSRSDVDSTTATPTPPPTSDDYKTKQATAYDTIFTSTS 126
Db 274 QAYKAAHK-----ABEEKAKTQRLITTESENINLKKRP----- 307
QY 127 LADIAALVSLQDAVN-----IKDTAATDE-----ETAIAAEWETKNADAIAKVGAI 174
Db 308 -----NDVSNRDKKKNSETAKTDEVEKQRAAEAAKAVETEKQRA-----AEA 350
QY 175 TELAKYASNDQAILDSGLKLTSPDLLQTLALQSVANNKAAELLKEMQDNPVPGKTPAI 234
Db 351 TKVAE-----AEKRAAEAAKAVETEKQRAAEATKV 381
QY 235 A-----QSLVDQTDATATQIEKDGNAIGDAYFAGNAGAVENAKSN-----NSISNID 283
Db 382 AEAEKQKAAEAARAVETEKQRAAEATKVAAEKQRAAEAMKVAEAEKQKAAEAATKVAAE 441
QY 284 SAKAAIATAKTOIAEAK-KFPDPSPILOQAEOM-----VIOAEKDLNKNIPADGSDVP 335
Db 442 KQKAAEA---TKVAEAEKQKAAEAATKVAAEKQKAAEAATKVAAEKQ---KAAEAATKV 494
QY 336 NPGTTVGSGKQSSIGSIRVSMILLDDAENETASILMSGFRQMIHMFNTEPNPSQAAQOE 395
Db 495 E-----AEKQKAAEAATKV-----EAEKQKAAEAATK-----VAEAEKQKAAEAAT 533
QY 396 LAAQARAARAAAGDDSSAAAALADAQKALEAALGKAGQOQGI 436
Db 534 KVAEAEKQKAA--EATKVAAEAEKQKAAEAARAKAMESQKORFL 572

```

RESULT 3

```

US-09-115-746-8
; Sequence 8, Application US/09115746
; Patent No. 6228601
; GENERAL INFORMATION:
; APPLICANT: Kirchhoff, Louis V.
; APPLICANT: Otsu, Keiko
; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington, D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/115,746
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/216,894
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: BENT, Stephen A.
; REGISTRATION NUMBER: 29,768
; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202)672-5300
; TELEFAX: (202)672-5399

```

TELEX: 904136
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 643 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-8

Query Match 6.2%; Score 150.5; DB 4; Length 643;
Best Local Similarity 21.5%; Pred. No. 0.0015;
Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;

QY 7 PGIDETERTPPADLSAAGLEASAAKSAEAOIAGAEAPKESKTSVERWSTLRSVAVN 66
DB 216 PPKSDVLRGSPSOL--QOENNTITNSKKWKLEKVKKAKEKLDINRATKLEERN 273
QY 67 ALMSLADKLGIASSSSSTSRSDVDTTATPTPPPTSDDYKTOAQATYDTFTSTS 126
DB 274 QAYKAAHK--AEBEKAKTFORLITFESINLKKRP----- 307
QY 127 LADTQAAVLSQDAVNTN-----IKDTAATDE-----ETAIAAEWETKNADAIVGAQI 174
DB 308 -----NDAVSNRDKKNSETAKTDEVEKORAAEAAKAVETEKORA-----AEA 350
QY 175 TELAKYASDNQAILDLSGLKLTSPDLLQTLQSVANNKKAELKEMODNPVPGKTPAI 234
DB 351 TKVAE-----AEKKAEEAAKAVETEKORAAEATKV 381
QY 235 A-----QSLVDQTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSN-----NSISNID 283
DB 382 AEAEKQKAAEAAKAVETEKORAAEATKVAAEAKQRAEAEMKVAEAEKQAAEATKVAAEAE 441
QY 284 SAKAAIATAQTQIAEAK-KFPDSPILQEAQM-----VIOAEKDLKNIKPADGSDVP 335
DB 442 KQKAAEA---TKVAEAEKQAAEATKVAAEAEKQKAAEATKVAAEAKQ-----KAAEATKVA 494
QY 336 NPGTGVGSKQGGSSIGSIRVSMLLDDAENETASILMSGFQMIHMFNTENPDSDQAAQOE 395
DB 495 E-----AEKQKAAEATKVA-----EAQKQAAEATK-----VAEAEKQKAAEAT 533
QY 396 LAAQARAARAGDSSAAALADAAQKALEALGKAGQOQGL 436
DB 534 KVAEAEKQKAA--EATKVAEAEKQKAAEAARAKAESQKQREL 572

RESULT 4
US-08-687-956A-23
Sequence 23, Application US/08687956A
Patent No. 5861157
GENERAL INFORMATION:
APPLICANT: BURNIE, JAMES P
APPLICANT: MATTHEWS, RUTH C
TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI
NUMBER OF SEQUENCES: 23
CORRESPONDENCE ADDRESSES:
ADDRESS: PILLSBURY, MADISON & SUTRO, LLP
STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
CITY: WASHINGTON
STATE: DC
COUNTRY: USA
ZIP: 20005
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/687,956A
FILING DATE: 29-JUL-1996

CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: GB 9401689.6
FILING DATE: 28-JAN-1994
ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N
REGISTRATION NUMBER: 16,773
REFERENCE/DOCKET NUMBER: 50885/222892
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202/861-3000
TELEFAX: 202/822-0944
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1566 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: peptide
HYPOTHETICAL: NO
ANTI-SENSE: NO
FRAGMENT TYPE: internal
ORIGINAL SOURCE:
ORGANISM: Streptococcus sobrinus
US-08-687-956A-23

Query Match 6.1%; Score 147; DB 2; Length 1566;
Best Local Similarity 22.6%; Pred. No. 0.0097;
Matches 106; Conservative 75; Mismatches 209; Indels 80; Gaps 19;
QY 48 KESKTSVERWSTLRSVAVNALSADKLGIASSSSSSSTS--RSADVSTTATAPTPTPP 105
DB 10 RKSKISRTLCGALLGTAL--LASVTGOKALAEETSTSTSGVNTAVVGTGTGNATNLPD 67
QY 106 TSDYKTOAQATY-----DTFTSTSLADIAALVSLQDAVNTNIDTAATDEATAA 158
DB 68 KQDNPSSQAEETSOAAGQKGTGAMSDVSTSELDEAAKSAQAEAGVTVSQDATVDRGTVETS 127
QY 159 EWEETKNADAIK--VGAOITELAKYASD-----NOATLDSGLKLTSPDLLQTLQSV 209
DB 128 DEANQKETEIKDDYSKQAADIQKTTEDYKAAVARNQAEETD---RITQENAAKKAQYEGDL 184
QY 210 NNKAAELLKEMODNPVVPKTPAIAQSLVDOTDATATQIEKDGNAIGDAYFAGQNASGA 269
DB 185 AAKK-AEVERTITNEN-----AQKADY-EAKLAQYOKDLAAVQOANDSOAAVAA 232
QY 270 VEHAKSNNSISNIDSAKAAIATAKTQIAEA-----QKKFDPSPILQ-----EAE 313
DB 233 AKEA-YDKELARVQOANAA---AKKEYEALAAANTTNEQIKAEENATQORNAQAKADYE 288
QY 314 QMVIQAEKDLKNIKPADGSDVPNPGTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMS 373
DB 289 AKLAQYKDLAAQAQSGNATNEADYQAKKAAVEQELARV-----QAANAANK---Q 335
QY 374 GFQMIHMFNTENPDSDQAQOELAAQAPAKAAAGDDSSAAALADAAQKALEALGKAGQOQ 433
DB 336 AYEQAALAAANTAKN--AQITAEENAIQORNAQAKA--NYEAKLAQYQKDLAAA-----QS 385
QY 434 GILNALGOIASAAVVSAGVLPLOQVLMTRARYQAYVEQKLISEEDLSAV 483
DB 386 G--NAANEADYQEKLAAYERELARVQOANAAAKOEYKQVQOANAKNAEI 433
RESULT 5
5352450-2
Patent No. 5352450
APPLICANT: KOGA, TOSHIHIKO; OKAHASHI, NOBUO; TAKAHASHI, ICHIRO;
SHIBUYA, KOJI; OHTA, HIROAKA
TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL
CARRIES AND VACCINAL COMPOSITIONS FOR DENTAL CARRIES USED AS NASAL
DROP
NUMBER OF SEQUENCES: 2
CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/07/529,602
FILING DATE: 29-MAY-1990
SEQ ID NO: 2
LENGTH: 1565
5352450-2

Query Match 6.1%; Score 146.5; DB 6; Length 1565;
Best Local Similarity 21.8%; Pred. No. 0.01;
Matches 101; Conservative 81; Mismatches 190; Indels 91; Gaps 21;

QY 48 KESKTSVERWSILRSVAVNALMSLADKLGIASSN--SSSTSRSDVDSTTATATPPPP 105
DB 2 KVKTYGFRKSKIKLGGVTVAAVSVAGQKVFADETTTSDVDKVVGTQTGNPA 61
QY 106 T-----SDYKTOQATAYDTFTST-----SLADIQAALVS-LQDAVNIKD 146
DB 62 TQLEAQSASKEAEQSTKLERQMVHTIEVPKTDLDQAADAKSAGVNVVQADVN-KG 120
QY 147 TAATDETAIAEWE-----TKNDAIK-----VGAQITELAKYASDNQAILSLG 192
DB 121 TVKTPPE-AVGKETEIKEDYTKQAEIKTKTDQTKSDVAHEAEVAKIKAKNQATKEQYE 179
QY 193 KLTSDLLQALQSVANNKAA--ELLKEMQDNVVPVGPATPAQSLVDQTDATATQIE 250
DB 180 KMAHKAERIRINANAASKTAYEAKLAQYQADLAQVKTNNANQAAV-QKALAAVQAE 238
QY 251 ----KGNAGIDAYFAGQASGAVENAKSNNSISNIDSAAKATATAK-----TQIAEAQ-- 300
DB 239 LKRVQEAANAANKAAV---DTAVANAANKTEIAAANEEIRKNATAKAAYETKLAQYQAE 295
QY 301 -KFPDPSILQAE-QWVIOA-EKDLNIRPAD-----GSDVNPVGTGGS- 344
DB 296 LKRVQEAANAANEADYQAKLTAYOTELARVOKANADAKATYEAVAANNAKNAALTAE 355
QY 345 -----KQGSIGSIRVSMILLDDAENETASILSGFRQIMHMENTNP 388
DB 356 IKERNENAKATYEAALQYEAADLAQVKKANAANEADYQAK---LTAYOTELARVOKANAD 412
QY 389 SOAQOELAAQARAQAA--GDSAAAL-ADAQKALEALGK 428
DB 413 AKAAEYAAVAANAANAALTAENTAENTAIKRNADAKADYEAELK 455

RESULT 6
US-08-216-894-2
Sequence 2, Application US/08216894
Patent No. 5876734
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/216.894
FILING DATE: 24-MAR-1994
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELECOMMUNICATION INFORMATION:

TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-216-894-2

Query Match 5.7%; Score 138.5; DB 2; Length 564;
Best Local Similarity 21.4%; Pred. No. 0.01;
Matches 97; Conservative 54; Mismatches 165; Indels 137; Gaps 17;

QY 7 PGIDETERTPADLSAQGLEASAAKSAABORIAGAEAKPKESKTSVERWSILRSVAVN 66
DB 216 PKKSDLVPRGSPSQL--QQAENNITNSKKEMTKLREKVKKAKEKLDAINRATKLEERN 273
QY 67 ALMSLADKLGIASSNSSSTSRSDVDSTTATATPPPTSDDYKTOQATAYDTFTSTS 126
DB 274 QAYKAAHK---ABEEKAKTFQRLITFESINILKKRP----- 307
QY 127 LADIQAALVSLQDAVTN-----IKDTAATDE-----ETAIAAEWETKNADAIVQAQI 174
DB 308 -----NDAVSNRDKKKNSETAKTDEVEKQRAAEAAKAVETEKQRA-----AEA 350
QY 175 TELAKYASDNQAILDSGLKLTSDLLQALQSVANNKAAELLKEMQDNVVPVGPATPAI 234
DB 351 TKVAE-----AEKRAAEAAKAVETEKQRAAEATKV 381
QY 235 A---OSLVDQTDATATQIEKDGNAIGDAYFAGQASGAVENAKSNNSISNIDSAAKAAIA 290
DB 382 ABEAKQKAAEAQKAVETEKQRAAEATKVAAEKQRAEAMK-----VAEAEKQKAAEA 434
QY 291 TAKTOIAEAQK-KFPDPSILQAEQM-----VIOAKDLKNIKPADSGSDVNPVGTG 342
DB 435 ---TKVAEAEKQKAAEAATKVAAEKQKAAEAATKVAAEKO---KAAEATKVAE----- 481
QY 343 GSKOQSGSIGSIRVSMILLDDAENETASILSGFRQIMHMENTNPDSQAAQELAAQARA 402
DB 482 AEKQKAAEAATKV-----AEKQKAAEATK-----VAEAEKQKAAEAATKVAAEAK 526
QY 403 AKAAAGDDSAANA-----LADAQKALEALGKAGQ 431
DB 527 QKAAEAATKVAAEKQKAAEAATKVAAEKQKAGE 559

RESULT 7
US-09-115-746-2
Sequence 2, Application US/09115746
Patent No. 6228601
GENERAL INFORMATION:
APPLICANT: Kirchoff, Louis V.
APPLICANT: Otsu, Keiko
TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
TITLE OF INVENTION: WITH TRYPANOSOMA CRUZI
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 3000 K Street, N.W., Suite 500
CITY: Washington, D.C.
COUNTRY: USA
ZIP: 20007-5109
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/115,746
FILING DATE:

Db 2031 I-----DSEDDLQECISS-----AMPKKKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2079

QY 416 ADAQKA-LEAALG-----KAGQQ--QGILNALGQIASAAVVS 449
 Db 2080 KDQRPDSEHGLSPDSENFDMKALQEGANSIVSSLHQAAAAACLS 2124

RESULT 9

US-08-289-548A-7
 ; Sequence 7, Application US/08289548A
 ; Patent No. 5648212

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25

CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/289,548A
 FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.46943
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100
 TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 ORIGINAL SOURCE:
 ORGANISM: Homo sapiens
 IMMEDIATE SOURCE:
 CLONE: APC
 US-08-289-548A-7

Query Match 5.7%; Score 138; DB 1; Length 2842;
 Best Local Similarity 20.0%; Pred. No. 0.11;
 Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERPPADLSA-QGLESAANKSAEAGRIAGAEAKPE-----SKTDSV-----55

Db 1658 DLTIESPPELAGEVGRGAQSAGEFEKRDITPTIGRSTDEAQQGKTSSVTIPELDDNKA 1717

QY 56 ERWSILERSAVNALM-----SLADKLGIASSNSSSTSRSDVDSTATATPTPP 103

Db 1718 EEGDILAEINCINSAMPKGSKHKPRVKKIMDQVOQAASSAPNKNQLDGKKKPTSPVKP 1777

QY 104 PPTSDDYKTAQATYDT-----IFT-----STSLADIAQALVSLQDAV----141
 Db 1778 IPONTEYTRVRKNADSKNNLNAERVFSNKKONLNKNSKDFNDKLPNNEDRVGSGF 1837
 QY 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191
 Db 1838 AFDSPPHYPIEGTPYCFSRNDSLSLDFDDDDVDLSREKAEL-----RRAKENK---ESE 1890
 QY 192 GKLTSEDLLOTTALLOSANNKAAELLKEM-----QNPVVVPGKTPAIAOSLVDTATAT 247
 Db 1891 AKVTS-----HTELTSNQOSANKTOAIAKQPINRGQPKILQ-KOSTFPQSSKDIPRGAA 1945
 QY 248 QIEKDNGAIGDAYFAGONASGAVENAKSNNSISNIDSARAAAIATATQIAEAKKKFPDSP 307
 Db 1946 TDEKLQN-----FAIENTPVCFSHNSSUSLSLSDID-----1975
 QY 308 ILQEAQOMVIAEKDLKNIPKADGSDVPNPGTTVGSGKQ-----QGSSIGSIR 355
 Db 1976 --QENNN--KENEPKETEPDPSQGEPSKQASGVAPKSFHVEDTPVCFSRNSSLSSLS 2030
 QY 356 VSMLLDDAENETASILMSGFROHMINTENPDQSQAAQOELAAQAAKAAAGDSDAAAL 415
 Db 2031 I-----DSEDDLQECISS-----AMPKKKKPSRLKGDNEKHSRNMGGILGED-LTLDL 2079
 QY 416 ADAQKA-LEAALG-----KAGQQ--QGILNALGQIASAAVVS 449
 Db 2080 KDQRPDSEHGLSPDSENFDMKALQEGANSIVSSLHQAAAAACLS 2124

RESULT 10

US-08-452-654-7
 ; Sequence 7, Application US/08452654
 ; Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
 APPLICANT: ANAND, RAKESH
 APPLICANT: CARLSON, MARY
 APPLICANT: GRODEN, JOANNA
 APPLICANT: HEDGE, PHILIP J.
 APPLICANT: JOSLYN, GEOFF
 APPLICANT: KINZLER, KENNETH
 APPLICANT: MARKHAM, ALEXANDER F.
 APPLICANT: NAKAMURA, YUSUKE
 APPLICANT: THLIVERIS, ANDREW
 TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
 TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie & Beckett
 STREET: 1001 G Street, NW
 CITY: Washington
 STATE: D.C.
 COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patent in Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/452,654
 FILING DATE: 25-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940
 FILING DATE: 08-AUG-1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Kagan, Sarah A.
 REGISTRATION NUMBER: 32,141
 REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-08-452-654-7

Query Match 5.7%; Score 138; DB 1; Length 2842;
Best Local Similarity 20.0%; Pred. No. 0.11; Indels 144; Gaps 23;
Matches 105; Conservative 82; Mismatches 194;

Qy 11 DETERTPPADLSA-QGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55
Db 1658 DLTIESPNNELAAGGVGGAQSGFEKRDITPTGRSTDEAOGKTSSTVIPELDDNKA 1717
Qy 56 ERWSILRSVAVNLM-----SLADKLGIASSNSSSTSRSDVSTTATATPPP 103
Db 1718 EEGDILAEICINSAMPKSKHPRVKIKMDVQQAASASSAPNKLQDGGKKKPTSPVKP 1777
Qy 104 PPTSDDYKTOAQAYDT-----IFT-----STSLADIQAALVSLQDAV----- 141
Db 1778 IPQNTFYRTRVKNADSKNNLNAERVFSDNKKQKNNKSKDFNDKLPNNEDKRVGSF 1837
Qy 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191
Db 1838 AFDSPPHVTPIEGTPYCFSRNDSLSDLFDDDDVDSLREKAEL-----RKAKENK---ESE 1890
Qy 192 KGLTSFDLLQTLALLOSANNKAAELLEM-----QDNVVPVPGKTPAIAQSLVDQTDATAT 247
Db 1891 AKVTS-----HTELTSNQQSANKTOAIAKQPINRGPKPILO-KQSTFPQSSKDIPDRGAA 1945
Qy 248 QLEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTQIAEAKKFPDSP 307
Db 1946 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1975
Qy 308 ILQEAQMVIOAEKDLNKKIPADGSDVPNPGTTVGGSKQ-----QGSSIGSIR 355
Db 1976 --QENNN---KENEPKETEPDQSGEPKPAQSGAPKSFHVEDTPVCFSRNSSLSSLS 2030
Qy 356 VSMLLDDAENETASILMSGFQMTMFNTENPDQQAQELAAQARAARAKAGDDSAAL 415
Db 2031 I-----DSEDDLQECISS-----AMPKKKPSRLKGNKHSRPNMGGIIGED-LTLDL 2079
Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALQIASAAVVS 449
Db 2080 KDIQRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLS 2124

RESULT 11
US-07-741-940-2
Sequence 2, Application US/07741940
Patent No. 5352775
GENERAL INFORMATION:
APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner, Birch, Mckie & Beckett
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/07741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: AMINO ACID
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-741-940-2

Query Match 5.7%; Score 138; DB 1; Length 2843;
Best Local Similarity 20.0%; Pred. No. 0.11; Indels 144; Gaps 23;
Matches 105; Conservative 82; Mismatches 194;

Qy 11 DETERTPPADLSA-QGLEASAAKSAEQAQRIAGAEAKPKE-----SKTDSV----- 55
Db 1659 DLTIESPNNELAAGGVGGAQSGFEKRDITPTGRSTDEAOGKTSSTVIPELDDNKA 1718
Qy 56 ERWSILRSVAVNLM-----SLADKLGIASSNSSSTSRSDVSTTATATPPP 103
Db 1719 EEGDILAEICINSAMPKSKHPRVKIKMDVQQAASASSAPNKLQDGGKKKPTSPVKP 1778
Qy 104 PPTSDDYKTOAQAYDT-----IFT-----STSLADIQAALVSLQDAV----- 141
Db 1779 IPQNTFYRTRVKNADSKNNLNAERVFSDNKKQKNNKSKDFNDKLPNNEDKRVGSF 1838
Qy 142 -----TNIKDT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDSL 191
Db 1839 AFDSPPHVTPIEGTPYCFSRNDSLSDLFDDDDVDSLREKAEL-----RKAKENK---ESE 1891
Qy 192 KGLTSFDLLQTLALLOSANNKAAELLEM-----QDNVVPVPGKTPAIAQSLVDQTDATAT 247
Db 1892 AKVTS-----HTELTSNQQSANKTOAIAKQPINRGPKPILO-KQSTFPQSSKDIPDRGAA 1946
Qy 248 QLEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKAIATAKTQIAEAKKFPDSP 307
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1976
Qy 308 ILQEAQMVIOAEKDLNKKIPADGSDVPNPGTTVGGSKQ-----QGSSIGSIR 355
Db 1977 --QENNN---KENEPKETEPDQSGEPKPAQSGAPKSFHVEDTPVCFSRNSSLSSLS 2031
Qy 356 VSMLLDDAENETASILMSGFQMTMFNTENPDQQAQELAAQARAARAKAGDDSAAL 415
Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKGNKHSRPNMGGIIGED-LTLDL 2080
Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALQIASAAVVS 449
Db 2081 KDIQRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAACLS 2125

RESULT 12

```
US-08-289-548A-2
; Sequence 2, Application US/08289548A
; Patent No. 5648212
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 102
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner & Allegretti, LTD
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/289,548A
; FILING DATE: 12-AUG-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.46943
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-289-548A-2

Query Match 5.7%; Score 138; DB 1; Length 2843;
Best Local Similarity 20.0%; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-QGLESAANKSAEAGRIAGAKPKPE-----SKTDSV----- 55
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1659 DLTIESPPNELAGEVGRGAQSGEKEKDTIPTGRSTDEAQQGKTSVTIPELDDNKA 1718
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 56 ERWSILRSVAVNALM-----SLADKLGIASSNSSSTSRSAVDSTTATAPPP 103
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1719 EEGDILAECINSAMPKSKHKPRVRKKIMDQVQQAASASSAPNKNLDGKKKKPTSPVKP 1778
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 104 PPTSDDKYKQAQYAYDT-----LFT-----STSLADIAQALVSLQDAV---- 141
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1779 IPQNTFYTRVRKNADSKNNLNARVFSNKKONLKNNSKDFNKLFPNNEDVRVGSF 1838
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 142 -----TNIKOT--AATDEETAIAAEWETKNADAIVKGAQITELAKYASDNOALDLSL 191
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1839 AFDSPHHYPIEGTFCFSRNDLSLDFDDDDVLSREKAEI-----RRAKENK-----ESE 1891
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 192 GKUTSFDLQTLAQSVANNKAAELLKEM-----QNPVVPKTPAIAQSLVDQTDATAT 247
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :

Db 1892 AKVTS-----HTELTSNQOSANKTQATAKOPINRGQPKPILO-KQSTFPOSSKDIPDRGAA 1946
QY 248 QIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATATQTAEAQKKEPDSP 307.
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1947 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1976
QY 308 ILQEAQMVIAQAKDLKNIKPADGSDVPNPGTTVGSKQ-----QGSSISGISIR 355
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1977 --QENNN--KENEPKETEPDPSQGEPSKQASGVAPKSFHYVEDTPVCFSRNSSLSSLS 2031
QY 356 VSMLLDDAENETASILMSFRQMIHMFNTENPDQSAQQAQELAAQARAAGAAGDDSAAL 415
| : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 2032 I-----DSEDDLQECISS-----AMPKKKPSRLKAGDNEKHSRNMGGTILGED-LTLDL 2080
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAVVS 449
Db 2081 KDITORPDSEHGLSPDSFNFQWKAIQEGANSIVSSLHQAAAAACLS 2125

RESULT 13
US-08-452-654-2
; Sequence 2, Application US/08452654
; Patent No. 5691454
; GENERAL INFORMATION:
; APPLICANT: ALBERTSEN, HANS
; APPLICANT: ANAND, RAKESH
; APPLICANT: CARLSON, MARY
; APPLICANT: GRODEN, JOANNA
; APPLICANT: HEDGE, PHILIP J.
; APPLICANT: JOSLYN, GEOFF
; APPLICANT: KINZLER, KENNETH
; APPLICANT: MARKHAM, ALEXANDER F.
; APPLICANT: NAKAMURA, YUSUKE
; APPLICANT: THLIVERIS, ANDREW
; TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
; TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
; NUMBER OF SEQUENCES: 94
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, NW
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20001-4598
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/452,654
; FILING DATE: 25-MAY-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/741,940
; FILING DATE: 08-AUG-1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Kagan, Sarah A.
; REGISTRATION NUMBER: 32,141
; REFERENCE/DOCKET NUMBER: 1107.035574
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2843 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-452-654-2

Query Match 5.7%; Score 138; DB 1; Length 2843;
```

Best Local Similarity 20.08; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

Qy 11 DETERTPADLSA-QGLEASAAKSAEQAORITAGAEAKPE-----SKTDSV----- 55
Db 1659 DLTIESPNELAAGEVGRGAQSGEFKRDITPTGRTDEAOGGKTSVTIPELDDNKA 1718
Qy 56 ERWSTLSRAVNALM-----SLADKLGIASSNSSSTSRSDVDSTTATATPP 103
Db 1719 EGGDILACINSAMPKSHKHPFRVKKIMDVOQASASSAPNKNQDGLPNNEDVRGSGF 1778
Qy 104 PPTSDYKTOAQATYDT-----IFT-----STLADIQAALVSLQDAV----- 141
Db 1779 IPQNTYRTRVRKNADSKNNLNAERVFSDNKKNSKDFNDKLPNNEDVRGSGF 1838
Qy 142 -----TNIKDT--AATDETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDSL 191
Db 1839 AFDSPHHYTPTEGTPYCFSRNDSLSLDFDDDDVLSREKAE-----RKAKENK---ESE 1891
Qy 192 GKLTSDLLQALLOSVAANNKAELLKEM-----QDNVPVPGKTPAIAQSLVDQTDATAT 247
Db 1892 AKVTS-----HTELTSNQOSANKTOAIAKQPINRGOPKPILO-KOSTFFQSSKDIPIRGAA 1946
Qy 248 QIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAIAATAKTAQIAEAQKFPDPS 307
Db 1947 TDEKLQ-----FAIENTPVCFSHNSLSLSDID-----QGSSIGSIR 355
Qy 308 ILQEAQMVIOAEKDKLNKIPADGSDVPNPGTTVGGSKQ-----QGSIGSIR 355
Db 1977 --QENNN--KENEPKETEPPDSOGEPKQASGYAPKSHVEDTPVCFSRNSSLSSLS 2031
Qy 356 VSMILDDAENETASTILMSGFROMIHMTENPDQAAQOELAAQAAKAGDSSAAAL 415
Db 2032 I-----DSEDDILQBCISS-----AMPKKKPSRLKGNKHSRPNMGGILGED-LTLDL 2080
Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQTASAAVVS 449
Db 2081 KDQIRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAAACL 2125

RESULT 14

US-08-452-655B-2
; Sequence 2, Application US/08452655B
; Patent No. 5783666

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA
APPLICANT: HEDGE, PHILIP J.
APPLICANT: JOSLYN, GEOFF
APPLICANT: KINZLER, KENNETH
APPLICANT: MARKHAM, ALEXANDER F.
APPLICANT: NAKAMURA, YUSUKE
APPLICANT: THLIVERIS, ANDREW
TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC
TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS
NUMBER OF SEQUENCES: 102
CORRESPONDENCE ADDRESS:
ADDRESSEE: Banner & Wilcoff, Ltd.
STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/452.655B
FILING DATE: 25-MAY-1995

CLASSIFICATION: 530
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/289,548
FILING DATE: 12-AUG-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 07/741,940
FILING DATE: 08-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.49964
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2843 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-452-655B-2

Query Match 5.78; Score 138; DB 1; Length 2843;
Best Local Similarity 20.08; Pred. No. 0.11;
Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

Qy 11 DETERTPADLSA-QGLEASAAKSAEQAORITAGAEAKPE-----SKTDSV----- 55
Db 1659 DLTIESPNELAAGEVGRGAQSGEFKRDITPTGRTDEAOGGKTSVTIPELDDNKA 1718
Qy 56 ERWSTLSRAVNALM-----SLADKLGIASSNSSSTSRSDVDSTTATATPP 103
Db 1719 EGGDILACINSAMPKSHKHPFRVKKIMDVOQASASSAPNKNQDGLPNNEDVRGSGF 1778
Qy 104 PPTSDYKTOAQATYDT-----IFT-----STLADIQAALVSLQDAV----- 141
Db 1779 IPQNTYRTRVRKNADSKNNLNAERVFSDNKKNSKDFNDKLPNNEDVRGSGF 1838
Qy 142 -----TNIKDT--AATDETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDSL 191
Db 1839 AFDSPHHYTPTEGTPYCFSRNDSLSLDFDDDDVLSREKAE-----RKAKENK---ESE 1891
Qy 192 GKLTSDLLQALLOSVAANNKAELLKEM-----QDNVPVPGKTPAIAQSLVDQTDATAT 247
Db 1892 AKVTS-----HTELTSNQOSANKTOAIAKQPINRGOPKPILO-KOSTFFQSSKDIPIRGAA 1946
Qy 248 QIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAIAATAKTAQIAEAQKFPDPS 307
Db 1947 TDEKLQ-----FAIENTPVCFSHNSLSLSDID-----QGSSIGSIR 355
Qy 308 ILQEAQMVIOAEKDKLNKIPADGSDVPNPGTTVGGSKQ-----QGSIGSIR 355
Db 1977 --QENNN--KENEPKETEPPDSOGEPKQASGYAPKSHVEDTPVCFSRNSSLSSLS 2031
Qy 356 VSMILDDAENETASTILMSGFROMIHMTENPDQAAQOELAAQAAKAGDSSAAAL 415
Db 2032 I-----DSEDDILQBCISS-----AMPKKKPSRLKGNKHSRPNMGGILGED-LTLDL 2080
Qy 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQTASAAVVS 449
Db 2081 KDQIRPDSEHGLSPDSENFDMKATQEGANSIVSSLHQAAAAAACL 2125

RESULT 15

US-08-452-655B-7
; Sequence 7, Application US/08452655B
; Patent No. 5783666

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS
APPLICANT: ANAND, RAKESH
APPLICANT: CARLSON, MARY
APPLICANT: GRODEN, JOANNA

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:38:03 ; Search time 96.2 Seconds
(without alignments)
387.999 Million cell updates/sec

Title: US-09-391-606-8

Perfect score: 2412

Sequence: 1 MNPICGPGPIDETERTPPAD.....QKLISEDLNSAVDHHHHH 490

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 219241 seqs, 76174552 residues

Total number of hits satisfying chosen parameters: 219241

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR_68.*

1: pir1.*

2: pir2.*

3: pir3.*

4: pir4.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2214	91.8	715	2 I40729	hypothetical 76k p
2	2188	90.7	651	2 D72042	conserved hypothet
3	2188	90.7	651	2 E6581	CHLPN 76 kDa homol
4	451.5	18.7	647	2 G71490	hypothetical prote
5	201	8.3	973	2 C85693	probable membrane
6	190	7.9	1122	2 G44887	probable tail fibe
7	188	7.8	2055	2 T31110	extracellular matr
8	178	7.4	1822	2 S3441	EF protein - strep
9	174.5	7.2	545	2 E84327	Htr7 transducer [i
10	174.5	7.2	545	2 T46811	halobacterial tran
11	171	7.1	627	2 F84194	Htr14 transducer [
12	170	7.0	1128	2 T30296	R27-2 protein - Tr
13	166.5	6.9	2232	2 T34434	hypothetical prote
14	160.5	6.7	544	2 T44938	transducer protein
15	160.5	6.7	1156	2 T34852	probable secreted
16	158	6.6	536	1 A47190	transducer protein
17	158	6.6	536	2 E84318	Htr1 transducer [i
18	158	6.6	810	2 T46810	halobacterial tran
19	158	6.6	810	2 F84327	Htr5 transducer [i
20	158	6.6	1365	2 T30822	Imp1 protein - Myc
21	156.5	6.5	641	2 C82206	methyl-accepting c
22	156	6.5	990	2 T51618	nucleolar phosphop
23	155.5	6.4	582	2 S24545	intermediate filam
24	155	6.4	892	2 T50985	related to transcr
25	155	6.4	1147	2 T35781	hypothetical prote
26	154.5	6.4	764	2 A84328	Htr2 transducer [i
27	154.5	6.4	765	1 T44946	transducer protein
28	154.5	6.4	1262	2 T22523	hypothetical prote
29	153	6.3	642	1 T44253	transducer protein

30	152.5	6.3	860	2 T14650	tail fiber protein
31	152.5	6.3	978	2 T14968	phage lambda-relat
32	152.5	6.3	1528	2 A60338	surface antigen A
33	152	6.3	1561	1 S06839	surface antigen sp
34	152	6.3	3488	2 T34418	hypothetical prote
35	151.5	6.3	1238	2 T03465	probable exonuclea
36	150.5	6.2	643	2 H84305	Htr8 transducer [i
37	149	6.2	1566	2 A43607	cell surface antig
38	148	6.1	571	2 D86164	hypothetical prote
39	148	6.1	729	2 E70803	hypothetical prote
40	148	6.1	881	2 S56032	probable membrane
41	148	6.1	1556	2 A60988	saliva-interacting
42	147.5	6.1	1302	1 JC6009	surface-located me
43	147	6.1	778	2 T48897	transducer protein
44	147	6.1	5327	2 T13564	microtubule-associ
45	146.5	6.1	446	2 H83098	hypothetical prote

ALIGNMENTS

RESULT 1

I40729

Hypothetical 76k protein - Chlamydomophila pneumoniae (strain AR39)

C:Species: Chlamydomophila pneumoniae, Chlamydia pneumoniae

C:Date: 16-Aug-1996 #sequence_revision 16-Aug-1996 #text_change 31-Mar-2000

C:Accession: I40729

R:Peréz-Melgosa, M.; Kuo, C.

Infect. Immun. 62, 880-886, 1994

A:Title: Isolation and characterization of a gene encoding a Chlamydia pneumoniae 76

A:Reference number: I40729; MUID:94156481

A:Accession: I40729

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-715 <R5>

A:Cross-references: GB:L23921; NID:g435961; PIDN:AAA23117.1; PID:g435962

A:Experimental source: strain AR-39

C:Comment: This is the hypothetical translation of a sequence that was reported as t

Query Match 91.8%; Score 2214; DB 2; Length 715;

Best Local Similarity 98.9%; Pred. No. 1.6e-110;

Matches 454; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY	1	MVNPICGPGPIDETERTPPADLSAQGLEASAAKSAEAOIRAGAEAKPKESKTSVERWSI	60
Db	257	LVPICGPGPIDETERTPPADLSAQGLEASAAKSAEAOIRAGAEAKPKESKTSVERWSI	316
QY	61	LRSAVNALMSLADKLGIASSNSSSTSRSDVDSTTATATPTPPPTSDDYKTAQATYDT	120
Db	317	LRSAVNALMSLADKLGIASSNSSSTSRSDVDSTTATATPTPPPTSDDYKTAQATYDT	376
QY	121	IFTSTSLADIQAALVSLQDAVNTIKDTATDEETAIAAEWETKNADAKVGAQITELAKY	180
Db	377	IFTSTSLADIQAALVSLQDAVNTIKDTATDEETAIAAEWETKNADAKVGAQITELAKY	436
QY	181	ASDNGAILDLSGLKTSFLLQALLOSVANNKKAELKEMODNPVPKTPFAIAQSLVD	240
Db	437	ASDNGAILDLSGLKTSFLLQALLOSVANNKKAELKEMODNPVPKTPFAIAQSLVD	496
QY	241	QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTAIAEAQ	300
Db	497	QTDATATQIEKDGNAIRDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTAIAEAQ	556
QY	301	KKFPSPILQEAQVIOAEKDLNKPAGDSVDNPNPCTTGGSKQOQSSIGSIRVSMLL	360
Db	557	KKFPSPILQEAQVIOAEKDLNKPAGDSVDNPNPCTTGGSKQOQSSIGSIRVSMLL	616
QY	361	DDAENETASILMSGFQMHFMENTENPDSSQAQOQLAAQAAKAAAGDDSAALADAQK	420
Db	617	DDAENETASILMSGFQMHFMENTENPDSSQAQOQLAAQAAKAAAGDDSAALADAQK	676
QY	421	ALAEAALGKAGQOQOQILNALGQIASAAVVSAGVLPLOQVL	459

```
|||||
Db 677 ALEAALGKAGQQGILNALGQIASAAVVSAGVLPLOQVL 715

RESULT 2
D72042
conserved hypothetical protein CP0018 [imported] - Chlamydophila pneumoniae (strains CWI
N:Alternate names: chlpn 76 kda homolog_1 (ct622); hypothetical protein CPn0728
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 23-Apr-1999 #sequence_revision 23-Apr-1999 #text_change 11-May-2000
C:Accession: D72042; D81623
R:Kalmán, S.; Mitchell, W.; Marathe, R.; Lammel, C.; Fan, J.; Olinger, L.; Grimwood, J.;
Nature Genet. 21, 385-389, 1999
A:Title: Comparative genomes of Chlamydia pneumoniae and C. trachomatis.
A:Reference number: A72000; MUID:99206606
A:Accession: D72042
A:Molecule type: DNA
A:Residues: 1-651 <ARN>
A:Cross-references: GB:AF001654; GB:AF001363; NID:g4377031; PIDN:AAD18867.1; PID:g437703
A:Experimental source: strain CWL029
R:Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,
C.; Dodson, R.; Gwinn, M.; Nelson, W.; DeBoy, R.; Kolonay, J.; McClarty, G.; Salzberg,
Nucleic Acids Res. 28, 1397-1406, 2000
A:Title: Genome sequences of Chlamydia trachomatis MoPn and Chlamydia pneumoniae AR39.
A:Reference number: A81500; MUID:20150255
A:Accession: D81623
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <ARN>
A:Cross-references: GB:AF002165; GB:AF002161; NID:g7188948; PIDN:AAF37914.1; PID:g718895
A:Experimental source: strain AR39, HL cells
C:Comment: This sequence was originally identified as homologous to part of a sequence (
PIR:H71490).
C:Genetics:
A:Gene: Cpn0728; CP0018

Query Match 90.7%; Score 2188; DB 2; Length 651;
Best Local Similarity 98.9%; Pred. No. 3.5e-109;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MYPNIGPGPIDETERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60
Db 1 MYPNIGPGPIDETERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60

QY 61 LRSVNALMSLADKLGIASSNSSSTSRSDVSTTATPTPPPTSDDYKTKQAQTAYDT 120
Db 61 LRSVNALMSLADKLGIASSNSSSTSRSDVSTTATPTPPPTSDDYKTKQAQTAYDT 120

QY 121 IFTSTSLADIQAALVSLQDAVNTIKDTAATDEETAAAEWETKNADAIKVGQAQITELAKY 180
Db 121 IFTSTSLADIQAALVSLQDAVNTIKDTAATDEETAAAEWETKNADAIKVGQAQITELAKY 180

QY 181 ASDNQAILDSLGKLSFLLQALQLQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240
Db 181 ASDNQAILDSLGKLSFLLQALQLQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240

QY 241 QTDATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSKAAATATAKTQIAEAQ 300
Db 241 QTDATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSKAAATATAKTQIAEAQ 300

QY 301 KFFPDSPILOEAQEWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360
Db 301 KFFPDSPILOEAQEWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360

QY 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQARAARAAAGDDSAALADAQK 420
Db 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQARAARAAAGDDSAALADAQK 420

QY 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVLP 454
Db 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVPP 454

RESULT 4
G71490
hypothetical protein CT622 - Chlamydia trachomatis (serotype D, strain UW3/Cx)
N:Alternate names: chlpn 76kda homolog CT622
C:Species: Chlamydia trachomatis
C:Date: 13-Sep-1998 #sequence_revision 13-Sep-1998 #text_change 08-Oct-1999
C:Accession: G71490
R:Stephens, R.S.; Kalmán, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitche
Science 282, 754-759, 1998
A:Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia
A:Reference number: A71570; MUID:99000809
A:Accession: G71490
A:Molecule type: DNA
A:Residues: 1-647 <ARN>
A:Cross-references: GB:AE001333; GB:AE001273; NID:g3329068; PIDN:AAG68226.1; PID:g33
```

```
RESULT 3
E86581
CHLPN 76 kda homolog_1 (CT622) [imported] - Chlamydophila pneumoniae (strain J138)
C:Species: Chlamydophila pneumoniae, Chlamydia pneumoniae
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 02-Mar-2001
C:Accession: E86581
R:Shirai, M.; Hirakawa, H.; Kimoto, M.; Tabuchi, M.; Kishi, F.; Ouchi, K.; Shiba, T.
Nucleic Acids Res. 28, 2311-2314, 2000
A:Title: Comparison of whole genome sequences of chlamydia pneumoniae J138.
A:Reference number: A86491; MUID:20330349
A:Accession: E86581
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-651 <STO>
A:Cross-references: GB:BA000008; NID:g8979100; PIDN:BAA98935.1; GSPDB:GN00142
A:Experimental source: strain J138
C:Genetics:
A:Gene: CPJ0728

Query Match 90.7%; Score 2188; DB 2; Length 651;
Best Local Similarity 98.9%; Pred. No. 3.5e-109;
Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MYPNIGPGPIDETERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60
Db 1 MYPNIGPGPIDETERTPPADLSAAGLEASAAKSAEQAIRIAGAEAKPKESKTSVERWSI 60

QY 61 LRSVNALMSLADKLGIASSNSSSTSRSDVSTTATPTPPPTSDDYKTKQAQTAYDT 120
Db 61 LRSVNALMSLADKLGIASSNSSSTSRSDVSTTATPTPPPTSDDYKTKQAQTAYDT 120

QY 121 IFTSTSLADIQAALVSLQDAVNTIKDTAATDEETAAAEWETKNADAIKVGQAQITELAKY 180
Db 121 IFTSTSLADIQAALVSLQDAVNTIKDTAATDEETAAAEWETKNADAIKVGQAQITELAKY 180

QY 181 ASDNQAILDSLGKLSFLLQALQLQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240
Db 181 ASDNQAILDSLGKLSFLLQALQLQSVANNKAAELKEMQDNVPVPGKTPAIQAQSLVD 240

QY 241 QTDATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSKAAATATAKTQIAEAQ 300
Db 241 QTDATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSKAAATATAKTQIAEAQ 300

QY 301 KFFPDSPILOEAQEWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360
Db 301 KFFPDSPILOEAQEWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMML 360

QY 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQARAARAAAGDDSAALADAQK 420
Db 361 DDENETASILMSGFQRMHMFNTENPDSSQAQOELAAQARAARAAAGDDSAALADAQK 420

QY 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVLP 454
Db 421 ALEAALGKAGQQGILNALGQIASAAVVSAGVPP 454
```

Qy	26	LESAANKSAEAOPIAGAEAKPKESKTDTSVERWLSRPSAVNALMSLADKLGIASSNSSS	85
		: : : : : : : : : : :	
Db	112	VEEVARNASAVAQNTAAAKKKSASDASTSAREANTHATDAADSARAASTASGOAAS-SAOS	170

•

QY	8	GPIDERTPPA----	DL	SAQLEASAANKSAE	AQ	RIAGAEAKPKSEK	SD	TSVERWSILRS	63
		:	:	:	:	:	:	:	:
		:	:	:	:	:	:	:	:
Db	93	GANTEDDARPEALRR	ELM---	VEEVARNASAVAQ	NTAAAKKSASDAS	TSAAREAATHAAD			149
		:	:	:	:	:	:	:	:
		:	:	:	:	:	:	:	:
QY	64	AVHALMSLADKLGI	ASSNSSSSTSR	SADVSDT	TATAPT	PPPTSDDYKTAQ	ATYDTFT		123
		:	:	:	:	:	:	:	:
		:	:	:	:	:	:	:	:

Db 150 AADSARAASSTAGQAAAS-SAGSASSAGTASTKATEASKSAAAAESSKSAASATAGAAKT 208
Qy 124 STSLADIQAALVSLQDAVNTKDTAATDETAIAAEWETKNADAIVKGAQITELAKYASD 183
Db 209 SETNAS-----ASLSAATSASTATTKASEAATSA-----RDAAASKEAKSSETNASSSA 259
Qy 184 NOAILDSGLKLTSPDLLQTLQSVANNKKAELLKEMQDNPVVPKTPATAOQLSVQD 243
Db 260 SSAASSATAAGNSAKAAKTS-----ETNARSSETAAGSASASAAAGSTKATAASSASAASTS 314
Qy 244 ATATQIEKDGNAIGDAYFAGONASGAVENAKSNNSISNIDSAAKAIATKATQIAEAQKKE 303
Db 315 A-----GQASASATAAGKSA-----ESAASSASTATTKAGEATEQ- 349
Qy 304 PDSPILOEAGQMVTOAEKDLNKNIPADGSDVPNPGTTVGGSKQOGSSIGSIRVSNMLDDA 363
Db 350 -----ASAAARSASAAKTSETNAKASSETSAESSKTAASASSASSASSASASASK--DEA 402
Qy 364 ENETASILMSGFQMIHMFNTENPDSQAQOELAAQA-----RAAKAAGDSDSAAALAD 417
Db 403 TRQASAAKSATTAATKATEAAGSATAAQAQSKSTAESAATRAETAAKRAEDIASVALED 462
Qy 418 AQKALEALGKAGQOQGLNALGQTASAAVYSAGVLPLOQVLWIRARYQAYVEOKLISEE 477
Db 463 ASTT-----KKGIV-----QLSSATNSTSETLAATP-----KAVKSAYDNAEKRLQK 504
Qy 478 DLNSA 482
Db 505 DONGA 509

RESULT 7
T31110
extracellular matrix binding protein - Abiotrophia defectiva (fragment)
C:Species: Abiotrophia defectiva
C:Date: 02-Sep-2000 #sequence_revision 02-Sep-2000 #text_change 02-Sep-2000
C:Accession: T31110
R:Manganelli, R.; van de Rij, I.
Infect. Immun. 67, 50-56, 1999
A:Title: Cloning and characterization of emb, a gene encoding the major adhesin of Streptococcus
A:Reference number: 420988; MUID:99081722
A:Accession: T31110
A:Status: preliminary;
A:Molecule type: DNA
A:Residues: 1-2055 <MAN>
A:Cross-references: EMBL:AF067776; NID:g3249002; PID:g3249003; PIDN:AAD03320.1
C:Genetics:
A:Gene: emb

Query Match 7.8%; Score 188; DB 2; Length 2055;
Best Local Similarity 23.9%; Pred. No. 0.034;
Matches 124; Conservative 68; Mismatches 197; Indels 130; Gaps 24;
Qy 19 ADLSAQGLEASAAKSAEQAQRIAGAEA-----KPKESKTSVERWSILRSVAVNALMSIAD 73
Db 1602 AKKAKDAIDAATSNADVTQAQDAGNAINAVPQTPTAKTD-----AKNAVTAQAD 1651
Qy 74 KLGIIASSNSSSTSRSDVSTTATPTPPPTSDDYKTAQATYDTFTSTSLADIAQ- 132
Db 1652 AKKDAIENDANIITREEKDAAKAKVDAEA-----TKAKNAID-----AATSNADVTAK 1698
Qy 133 -----ALVSLQDAVT-----NIKDTAATDEETAI-----AAEWETKNADAIVKGAQITE 176
Db 1699 QNEGTAKINDVPQTPTAKTAKNAVDQAATDKKSAIENDPALTRREEKDAAKAVDAEATK 1758
Qy 177 LAKYASDNOAILDSGLKLTSEDLLOTALLQSVANNKKAELLKEMQDNPV--VPKTPAI 234
Db 1759 -AKNAID-----AATSNADVTQAQDAGNAINAVP--QTPTA 1792
Qy 235 ---AQSLVDQ-----TDATATQIEKD--GNAIGDAYFAGONASGAVENAKSNNSI- 279
Db 1793 KTDKNAVDQAATDKKAAIENDPALTRREEKDAAKAVDA--EAKKAKDAIDAATSNADVT 1850

Qy 280 SNIDSAAKAI-----ATAKTQIAEA-----QKFPDPSPILOEAGQMVIOAEKDLKN 325
Db 1851 AQKDAGKDAINAVPQTPTAKTAKNAVDQAATDKKSAIENDPALTRREEKDAAKAVDAEA 1910
Qy 326 IKPADGSDVPNPGTTVGGSKQOGS-SIGSI-RVSNMLDDAENETASTILMSGFQMIHFN 383
Db 1911 KKAKDAIDAATSNADVTAKTQEGTQAINAVPQTPTAKTAKNAVDQAATDKKNAIENDPA 1970
Qy 384 TENPDSQAQOELAAQARAARAKAGDSDSAAALADAQKALEAALGKAGQOQ--ILNALGQ 441
Db 1971 LTRREEKDAAKAVDAEAKKAKDAID--AATSNADV-----TAKQNEGTAKINDVPQ 2019
Qy 442 IASAAVYSAGVLPLOQVLWIRARYQAYVEOKLISEEDLN 480
Db 2020 TPTAKTAKNAVD-----QAATDKKSAIENDAN 2047

RESULT 8
S33441
EF protein - Streptococcus suis
C:Species: Streptococcus suis
C:Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 15-Oct-1999
C:Accession: S33441
R:Smith, H.E.; Reek, F.H.; Vecht, U.; Gielkens, A.L.J.; Smits, M.A.
submitted to the EMBL Data Library, May 1993
A:Description: Repeats in an extracellular protein of wek-pathogenic strains are absent
A:Reference number: S33441
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1822 <SMI>
A:Cross-references: EMBL:X71880; NID:g298031; PIDN:CAA50714.1; PID:g298032

Query Match 7.4%; Score 178; DB 2; Length 1822;
Best Local Similarity 22.9%; Pred. No. 0.098;
Matches 122; Conservative 80; Mismatches 201; Indels 130; Gaps 23;
Qy 13 TERPPADLSAQGLEASAAKSAEQAQRIAGAEKPKES-KTDSVERWSIL-RSVAVALMS 70
Db 1201 TEKAKAAELAGEKSLTDTGKEARDAVELAKDELAKAEKTEEEETKIVLEKLAEDTRKA 1260
Qy 71 LADKLGIIASSNSSSTSRSD-VDSSTTATPTPPPTSDDYKTAQATYDTFTSTSLAD 129
Db 1261 IEDPNLSDEKQAEIKKLTDAVAKTLATI-----RDNAKRTQEAKEA-----QALAD 1309
Qy 130 IQAALVSLQDAVTNIKDTAATDEETAI--AEWETKNADAIVKGAQITELAKYASDQAI 187
Db 1310 LEKA-----KETQKIADKAAIDRLTILVKDGELEATKQDAKKNKIAKAAAAKEAIAISPN 1364
Qy 188 LDSLGKLTSPDLLQTLQSVANNKA-----AELLKEMQDNPVVPKTPATAQSLVD 240
Db 1365 LTDEAKTKFTD-----AVDAEAKANDAIASAATSPADVOKE-EDAGVA-----ATAEOVL 1414
Qy 241 QTDATA-TQIEKDGNAIGDAYFAGONASGA-----VENAKSNNSISNIDSAAKAI 289
Db 1415 AAKQDAKKNIAKAAAAKEAIGSNPLNTDAEKKFTTDAVDAEAKANDAI- 1465
Qy 290 ATAKTQIAEAQKFPDPSPILOEAGQMVTOAEKDLNKNIPADGSDV-----PNFGTT- 340
Db 1466 --AATSPADVOKE-EDAGVAAIAEDVLDAKQDAKKN-IAKESDAAKSAIDANPLNTDAE 1521
Qy 341 -----VGGSKQOGSSIGSIRVSM- -LDDAENETASTILM 372
Db 1522 KESAKKAVDAKAAKTAATDAIDASTSPVEAQSAEDKGVGIAQDVLDAKQDAKKNIAKEVA 1581
Qy 373 SGFQMIHMFNTENPDSQAQOELAAQARAARAKAGDSDSAAALADAQKALEAALGKAGQ 432
Db 1582 AAKEAIDANPLNSDAEKEASKAVDADAKATTTDAID--ASTSPVEAQSAEDKGVGSI--R 1637
Qy 433 QGILNALGQIASAAVVSAGVLPLOQVLWIRARYQAYVEOKLISEED-LNSAVD 484

100

0
1
2
3
4
5
6
7
8
9
A
B
C
D
E
F
G
H
I
J
K
L
M
N
O
P
Q
R
S
T
U
V
W
X
Y
Z
[
\
]
^
_
`
a
b
c
d
e
f
g
h
i
j
k
l
m
n
o
p
q
r
s
t
u
v
w
x
y
z
{
|
}
~
_

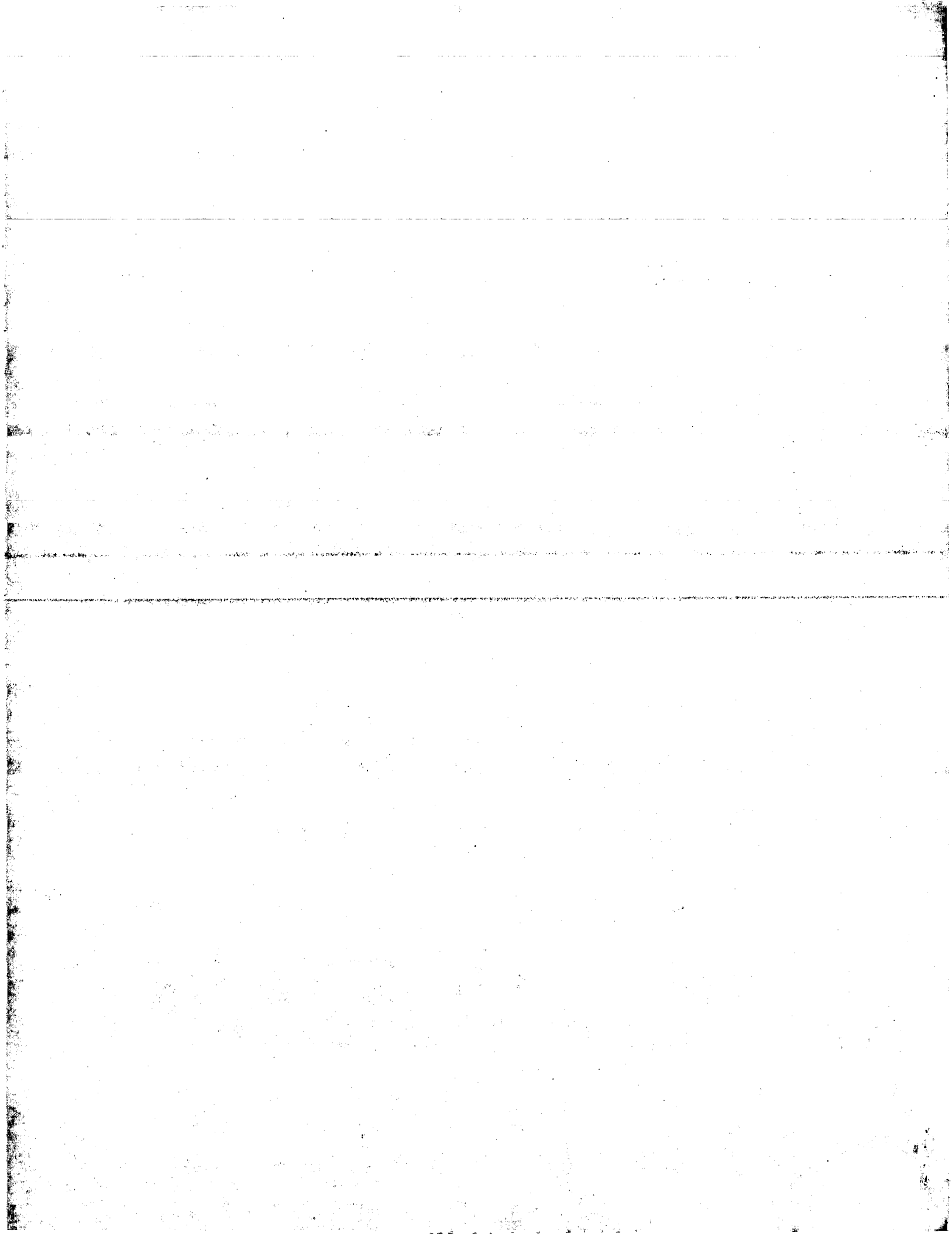
	Query Match	7.0%;	Score 170;	DB 2;	Length 1128;
	Best Local Similarity	22.9%;	Pred. No. 0.14;		
	Matches 117;	Conservative 82;	Mismatches 222;	Indels 90;	Gaps 20;
QY	10 IDERTPTPADLS----	AAGLEASAAKSAEAO	RITAGAEA--KPKE	SKTDSVERWSILRSA	64
	: : :	: : :	: : :	: : :	:
Dd	614 VAEAEKKAABATKV	AEAEKQAEEATKV	AEAEKQAEEATKV	AEAEKQAEEATKV	673
	: : :	: : :	: : :	: : :	:
QY	65 VNALMSLADKLGI	ASSNSSSTSRSDV	STATTATPTPPPTT	SDDYTKQAOTYDTFTS	124
	: : :	: : :	: : :	: : :	:
Dd	674 EKQAAAEATKV	AEAEKQAEEATKV	AEAEKQAEEAT--	KVAEAEKQAEEA-----TK	725
	: : :	: : :	: : :	: : :	:
QY	125 TSLADIQAALVSLO	DATVNIKDRAATDE	EATAIAEWETKNADA	IKV-----GAQTTEL	177
	: : :	: : :	: : :	: : :	:
Dd	726 VAEAEKQAEEATKV	AEAE--KQAA--EAT	KVAEAEKQAEEATKV	AEAEKQAEEATKV	782
	: : :	: : :	: : :	: : :	:
QY	178 AKYASDNOAILD	SLGKITSFLLQTTAL	LOSIV--ANNKKAEBEL	LK-----EMODNPVPVGKT	231
	: : :	: : :	: : :	: : :	:

```

Query Match      6.98; Score 166.5; DB 2; Length 2232;
Best Local Similarity 19.7%; Pred. No. 0.52;
Matches 79; Conservative 70; Mismatches 174; Indels 79; Gaps 11;

Qy      4  PIGPGPIDETERTPPADLSAQ----GLEASAAANKSAEAQRIAGAEAKPKESKTUSVERWS 59
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      598 PSSOSPAPNTGSTTPTSSQSPSPSMNPSSSTPTGSSQSTITPEGSTASSPTGSGTGF 657
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      60  ILRGAVALMSL--ADKLGIASSNSSSTSRADVDSTATATAPTPPPPTSDDYTKTAQTA 117
      | : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db      658 SVATEVTSQSIVPGSSSLGCTGTHSSPSPSSLSSTSGMSTLTSEPPSS-----TQSGGA 713
      | : : : : : | | | | | | | | | | | | | | | | | | | | | |
Qy      118 YDTIFT-STSLADIQAALVSLQDAVTNIKTPAATDEETAIAAEWETKNADAIKVCAQITE 176
      | | | | : : : | | | | | | | | | | | | | | | | | | | | |
Db      714 QSTLTTPSPNPQSQTSSLESSTSGATTSSGSGAGTT-----MTSPSQSSSVG----- 759
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      177 LAKYASDNOAILDSLGKLTISFDLIQTALLOSANNKAAELLKEMQDNPPVPGKTPATAQ 236
      | | | | : : : | | | | | | | | | | | | | | | | | | | | |
Db      760 -SSQGSTSPAASTTSGENTSGSTQT-----PGSSVSTSA 793
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Qy      237 SLVDOTATATQIEKQNAIGDAYFAGQNASGA---VENAKSNNSISNIDSAKAAIATAK 293
      | | | | : : : | | | | | | | | | | | | | | | | | | | | |
Db      794 AILLSTQGSVSTNPGSTVTPSPVSGSTSSGSTVTVGSTEASTSGSSVASSPAPSTSQ 853
      | | | | : : : | | | | | | | | | | | | | | | | | | | | |
Qy      294 TQIAEAKKFPDPSILOEAEQWVIOAEKDLKNIPAGDSVDVNPFC-----TTVGGSKQQ 347
      | | | | : : : | | | | | | | | | | | | | | | | | | | | |
Db      854 N-----PNPSTSGSSMITOSPPYPSQSTSPVESSTTPSPSGPCTTITSTSPSPSO 903
      | | | | : : : | | | | | | | | | | | | | | | | | | | | |

```

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:42:39 ; Search time 76.51 Seconds
(without alignments)
234.816 Million cell updates/sec

Title: US-09-391-606-8
Perfect score: 2412
Sequence: 1 MVNPTGPGIDERTPPAD.....QKLISEEDLNSAVDHHHHH 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 100059 seqs, 36664827 residues

Total number of hits satisfying chosen parameters: 100059

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	190	7.9	1120	1 STFR_ECOLI	P76072 escherichia
2	174.5	7.2	545	1 HTR5_HALN1	Q48318 halobacteri
3	158	6.6	535	1 HTR4_HALN1	P33741 halobacteri
4	158	6.6	810	1 HTR4_HALSA	Q9hp84 halobacteri
5	158	6.6	641	1 HRPK_PSESY	Q48317 halobacteri
6	154.5	6.4	763	1 HTR2_HALN1	P41501 pseudomonas
7	154.5	6.4	763	1 HTR2_HALN1	Q9hp81 halobacteri
8	154.5	6.4	764	1 HTR2_HALSA	P71410 halobacteri
9	152.5	6.3	1528	1 SPAA_STRDO	P21979 streptococc
10	152	6.3	1561	1 SPAP_STRMU	P23504 streptococc
11	148	6.1	505	1 FLJB_SALTY	P52616 salmonella
12	148	6.1	881	1 YJH8_YEAST	P47033 saccharomyc
13	147	6.1	778	1 HTR6_HALSA	Q48319 halobacteri
14	146.5	6.1	774	1 STE_LAMBD	P03764 bacterioph
15	146	6.1	1714	1 SYEP_DROME	P28668 drosophila
16	145	6.0	705	1 CWBA_BACSU	Q02113 bacillus su
17	145	6.0	1565	1 PAC_STRMU	P11657 streptococc
18	144	6.0	1637	1 MRSP_STAAU	P80544 staphylococ
19	143.5	5.9	1306	1 MSB2_YEAST	P32334 saccharomyc
20	142.5	5.9	778	1 HTR6_HALN1	Q9hr92 halobacteri
21	140	5.8	436	1 Y868_CHLMU	O9pj11 chlamydia m
22	139.5	5.8	2492	1 TALA_DICDI	P54633 dictyostell
23	139	5.8	797	1 VGLX_HSVEB	P28968 equine herp
24	138	5.7	1609	1 LMGI_HUMAN	P11047 homo sapien
25	138	5.7	2843	1 APC_HUMAN	P25054 homo sapien
26	137.5	5.7	2453	1 NCRL_MOUSE	O60974 mus musculu
27	137	5.7	500	1 FLJB_SALAE	P52615 salmonella
28	137	5.7	535	1 HTR1_HALSA	P33955 halobacteri
29	137	5.7	1607	1 LMGI_MOUSE	P02468 mus musculu
30	136	5.6	1379	1 YFF9_SCHPO	O14066 schizosacch
31	136	5.6	2541	1 TAL1_MOUSE	P26039 mus musculu
32	136	5.6	2845	1 APC_MOUSE	O61315 mus musculu
33	135.5	5.6	986	1 GM13_RAT	Q62839 rattus norv

34	135.5	5.6	1061	1 TRC4_ECOLI	P27189 escherichia
35	135.5	5.6	1969	1 MYSA_CAEEL	P12844 caenorhabdi
36	135.5	5.6	2090	1 HFC1_MESAU	P51611 mesocricetu
37	134	5.6	564	1 M12_STRPY	P19401 streptococc
38	134	5.6	1036	1 NIT2_NEUCR	P19212 neurospora
39	134	5.6	2842	1 APC_RAT	P70478 rattus norv
40	133	5.5	1972	1 MYHB_HUMAN	P35749 homo sapien
41	132.5	5.5	573	1 YEYB_YEAST	P40095 saccharomyc
42	132.5	5.5	1500	1 SSP5_STRGN	P16952 streptococc
43	132	5.5	2349	1 TPR_HUMAN	P12270 homo sapien
44	131.5	5.5	439	1 Y579_CHLTR	O84583 chlamydia t
45	131.5	5.5	1065	1 SED4_YEAST	P25365 saccharomyc

ALIGNMENTS

RESULT	1
STFR_ECOLI	
ID	STFR_ECOLI
AC	P76072: P77560; STANDARD; PRT; 1120 AA.
DT	20-AUG-2001 (Rel. 40, Created)
DT	20-AUG-2001 (Rel. 40, Last sequence update)
DT	20-AUG-2001 (Rel. 40, Last annotation update)
DE	SIDE TAIL FIBER PROTEIN HOMOLOG FROM LAMBDOID PROPHAGE RAC.
GN	STFR OR B1372.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
{}1	
RN	SEQUENCE FROM N.A.
RC	STRAIN=K12 / MG1655;
RX	MEDLINE=97426617; PubMed=9278503;
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA	Mau B., Shao Y.;
RT	"The complete genome sequence of Escherichia coli K-12.";
RL	Science 277:1453-1474(1997).
RN	{2}
RP	SEQUENCE FROM N.A.
RC	STRAIN=K12;
RX	MEDLINE=97251357; PubMed=9097039;
RA	Alba H., Baba T., Fujita K., Kimura S., Kitakawa M., Kitagawa M.,
RA	Kasai H., Kashimoto K., Kimura S., Mori H., Mori T., Motomura K.,
RA	Makino K., Miki T., Mizobuchi K., Mori H., Mori T., Motomura K.,
RA	Nakade S., Nakamura Y., Nishimoto H., Nishio Y., Oshima T., Saito N.,
RA	Sampel G., Seki Y., Sivasubram S., Tagami H., Takeda J.,
RA	Takemoto K., Takeuchi Y., Wada C., Yamamoto Y., Horiuchi T.;
RT	"A 570-Kb DNA sequence of the Escherichia coli K-12 genome
RL	corresponding to the 28.0-40.1 min region on the linkage map.";
CC	DNA Res. 3:363-377(1996).
CC	!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
CC	-----
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration
CC	between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC	the European Bioinformatics Institute. There are no restrictions on its
CC	use by non-profit institutions as long as its content is in no way
CC	modified and this statement is not removed. Usage by and for commercial
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC	or send an email to license@isb-sib.ch).
CC	-----
DR	EMBL; AE0000234; AAC74454.1; ALT_INIT.
DR	EMBL; D90774; BAA14966.1; -
DR	EMBL; D90775; BAA14975.1; -
DR	HSSP; P04002; IWFA.
DR	Ecogene; EG13370; stfr.
DR	InterPro; IPR000122; Chemotaxis_transducer.
KW	Hypothetical protein; Fiber protein; Repeat; Complete proteome.
SEQ	SEQUENCE 1120 AA; 113779 MW; 542E59D71EE795B4 CRC64;

RESULT 2

ID	HTRS_HALN1	STANDARD;	PRT;	545 AA.
QC	Q48318; Q9HP85;			
AT	15-DEC-1998 (Rel. 37, Created)			
DC	15-DEC-1998 (Rel. 37, Last sequence update)			
DT	20-AUG-2001 (Rel. 40, Last annotation update)			
DD	HALOBACTERIAL TRANSDUCER PROTEIN V.			
DE	HTR7 OR HTPV OR YNG1759G.			
GN	Halobacterium sp. (strain NRC-1), and			
OS	Halobacterium salinarum..			
OC	Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;			
OX	Halobacterium.			
XX	NCBI_TaxID=64091, 2242;			
[1]	SEQUENCE FROM N.A.			
RN	STRAIN=NRC-1;			
RC	MEDLINE=20504483; PubMed=11016950;			
RA	Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,			
RA	Shukla H.D., Lasky S.R., Balliga N.S., Thorsson V., Sbrogna J.,			
RA	Swartzell S., Weir D., Hall J., Dahl T.A., Weltl R., Goo Y.A.,			
RA	Ledtkauser B., Keller K., Cruz R., Danson M.J., Hough D.W.,			
RA	Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,			
RA	Isenberger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,			
RA	Alem M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,			
RA	Ebbardt H., Lowe T.M., Liang P., Riley M., Hood L., DasSarma S.;			
RT	"Genomic sequence of Halobacterium species NRC-1."			
RL	Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).			
[2]	SEQUENCE FROM N.A.			
RN	STRAIN=S9;			
RC	SPECIES=H.salinarum; STRAIN=S9;			
RA	MEDLINE=96275896; PubMed=8674984;			

```

Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K., Oesterheilt D.; "A family of halobacterial transducer proteins"; FEMS Microbiol. Lett. 139:161-168(1996).
!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL TRANSDUCTION.
!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
-----
EMBL; AE005080; AAC19985.1; --
EMBL; X95589; CAA64842.1; --
InterPro; IPR000122; Chemotaxis_transducer.
InterPro; IPR003660; HAMPT.
Pfam; PF00015; MCPsignal; 1.
SMART; SM00304; HAMP; 1.
SMART; SM00283; MA; 1.
Transducer; Transmembrane; Complete proteome.
TRANSMEM 10 30 POTENTIAL.
TRANSMEM 44 64 POTENTIAL.
TRANSMEM 80 100 POTENTIAL.
SEQUENCE 545 AA; 57070 MW; FDD870389C2F428B CRC64;
-----
Query Match 7.2%; Score 174.5; DB 1; Length 545;
Best Local Similarity 21.7%; Pred. No. 0.028;
Matches 100; Conservative 79; Mismatches 192; Indels 89; Gaps 17;
QY 12 ETERTPADLSAGCLASANKSAEQAORIACAKPKESKTDVSERWSIIRSAVNALMSL 71
DB 118 EEERA-----EAERAREKAFAQAE--O-TAAESA-KODARERSAEIEQLAADLESQ 169
QY 72 ADKLGIASSNSSSTSRADVDSTTATAPPPPT-SDDYKTOAQATAYDPI-----FTST 125
DB 170 ATEVG-ATLEAASDGDLTARVDATTNAAETAVTVVNDMLTWRTIDEIQGFSTNVTT 228
QY 126 SLADIQAALVSLQDAVTNIKDT-----AATDEE-----TAIAEWETKNADAIVGAQITE 176
DB 229 ASREATAGAKEIQDASTVESQEIAAGTDQREQLESVAEEEMDSYSATVEEVAAATAQS 288
QY 177 LAKYASDNQAILDSLGLKT-----SFDLLQTALLQSVANNKAAELLKEMQDPNPVPGK 230
DB 289 VADTAADTTDVA-TAGKQTAEADAIDAIVQETMTQTVANVDALEDLTITIDD----- 340
QY 231 TPAIAGSLVDOTDAT-----ATQIEKGNAIGD--AYFAGONASGAVENAKSNNS 278
DB 341 ---IAELISDIACQTNWLANIEARACSGSGSGNGDFAVVADEVKELATESQRS AKD 397
QY 279 ISN-IDSKAAAIATAKTQIAEAOKKFPDSPILOEAEQMVIQAEKDKLNKIPADGSDVPNP 337
DB 398 IAELEEVSQSATTVVEEI-----RVAEQRVNDGAAA VEETVDFAGAVTENI 444.
QY 338 GTTVGSCKQGSSGISGRVSMILLDAENETASILMSCFROMIHMFNTENPDSSQAQBELA 397
DB 445 QETTDGVQE-----ISQAMDEQQRSEVVSS-----VVDIATISQATA 483
QY 398 AQARAKAAGDDSAALAADAQALFAALCAGQQQGINL 437
DB 484 DRANVSAASEEQ-TASITETVTSLSLAQAQDTLEDRLN 522
-----
RESULT 3
HTR1_HALN1 STANDARD; PRT; 535 AA.
ID HTR1_HALN1
AC P33741; Q9HPF6;
DT 01-FEB-1994 (Rel. 28, Created)
DT 01-JUN-1994 (Rel. 29, Last sequence update)

```

20-AUG-2001 (Rel. 40, Last annotation update)
SENSORY RHODOPSIN I TRANSDUCER (HTR-I) (METHYL-ACCEPTING PHOTOTAXIS
PROTEIN I) (MPP-I).

GN HTR1 OR HTR1 OR VNG1659G.
OS Halobacterium sp. (strain NRC-1), and
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091, 2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NRC-1;
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
RN [2]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 1-10; 350-377 AND 457-476.
RC SPECIES=H.halobium; STRAIN=FLX5R;
RX MEDLINE=93101637; PubMed=1465418;
RA Yao V.J., Spudich J.L.;
RT "Primary structure of an archaeobacterial transducer, a
RT methyl-accepting protein associated with sensory rhodopsin I.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:11915-11919(1992).
CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005075; AAG19913.1; -;
CC PIR; A47190; A47190.
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR000658; DUF5.
DR Pfam; PF00672; DUF5; 1.
DR SMART; SM0015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.
KW Transducer; Photoreceptor; Transmembrane; Methylation.
FT INT_MET 0
FT DOMAIN 1 13
FT TRANSMEM 14 28
FT DOMAIN 29 38
FT TRANSMEM 39 54
FT DOMAIN 55 535
FT MOD_RES 265 265
FT MOD_RES 272 272
FT MOD_RES 279 279
FT MOD_RES 463 463
FT MOD_RES 472 472
SQ SEQUENCE 535 AA; 56544 MW; B9945E4F66A9D091 CRC64;

Query Match 6.69; Score 158; DB 1; Length 535;
Best Local Similarity 18.79; Pred. No. 0.19;
Matches 88; Conservative 83; Mismatches 189; Indels 110; Gaps 15;

RESULT 4

HTR4_HALN1 STANDARD; PRT; 810 AA.
ID Q9HP84;
AC 20-AUG-2001 (Rel. 40, Created)
DT 20-AUG-2001 (Rel. 40, Last sequence update)
DE HALOBACTERIAL TRANSDUCER PROTEIN IV.
GN HTR5 OR HTRP4 OR VNG1760G.
OS Halobacterium sp. (strain NRC-1).
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=64091;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20504483; PubMed=11016950;
RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogna J.,
RA Swartzell S., Weir D., Hall J., Dahl T.A., Welter R., Goo Y.A.,
RA Leithauer B., Keller K., Cruz R., Danson M.J., Hough D.W.,
RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
RA Ehardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
RT "Genome sequence of Halobacterium species NRC-1.";
RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
CC -!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL
CC TRANSDUCTION (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; AE005080; AAG19986.1; -;
DR InterPro; IPR000122; Chemotaxis_transducer.

	CC	-!- FUNCTION: POTENTIALLY INVOLVED IN CHEMO- OR PHOTOTACTIC SIGNAL TRANSDUCTION.
	CC	-!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
	CC	-----
	CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
	CC	-----
	DR	EMBL: X95589; CAA64841.1;
	DR	InterPro: IPR000122; Chemotaxis_transducer.
	DR	InterPro: IPR000658; DUF5.
	DR	InterPro: IPR003660; HAMP.
	DR	Pfam: PF00672; DUF5; 1.
	DR	Pfam: PF00015; MCPsignal; 1.
	DR	SMART: SM00304; HAMP; 2.
	DR	SMART: SM00283; MA; 1.
	DR	Transducer; Transmembrane.
	KW	DOMAIN 1 38
	FT	TRANSMEM 39 59
	FT	DOMAIN 60 323
	FT	EXTRACELLULAR (POTENTIAL).
	FT	TRANSMEM 324 344
	FT	CYTOSOL (POTENTIAL).
	FT	DOMAIN 345 810
	SQ	SEQUENCE 810 AA; 85207 MW; 37B0F6046A39D9BA CRC64;
		Query Match 6.6%; Score 158; DB 1; Length 810;
		Best Local Similarity 21.8%; Pred. NO. 0.31;
		Matches 101; Conservative 75; Mismatches 169; Indels 118; Gaps 22;
	QY	27 EASAANKSA-EAQRAGAEAKPKESKTSVERWSILRSVAVNALMSLADKLGIASSNSSSS 85
	DB	398 EAEEAREQATEAQDAEAE---RERAEARERAEDAKADAEL-----AAELEAQ 444
	QY	86 TGRSADVSTTTATPTPPPTDDKYTKQAQTAYDTFTSLADIQAALVSLQDAVTNIK 145
	DB	445 AERYSDVMAACADGDLTRMPADDTDNEAMAAIAASF-NEMLAQWEHTIIDIOE----FA 499
	QY	146 DTAATDEETAIAEWETKNADAIVGAQIT-----ELAKYASDNQAILDLSL-GKLTSPDLL 200
	DB	500 DAVATASEEA-----EVGAADAERASGGVSESVQEIAGAADFQRNMLDVTSGEMT--DL- 551
	QY	201 QTALLQSVAANNKAELLKEMQDNVPVPGKTPAIQAQLVDQDATATQIEKDNGAIGDAY 260
	DB	552 -SAAIEVA-----ASDSVAESHQTA-ETARDGE-----580
	QY	261 FAGQNAGAVENAKSNNSINIDSAKAATATAKTQIAEAKKKFPD--SPILOEAQMVIQ 318
	DB	581 ---QTAEADAEI-RSLTVQEAIDATVQNVVEALDDQMAEI-SEIVDLISIDIAEQTNMLALN 634
	QY	319 AE-----KDKLNKPADGSDVPNPPTGVGSKOOGSIGSIIRYSVM 358
	DB	635 ANTEAARADKSGDFAVVADEVKDLAEETOESAGDIERRITEV--QSOTTATVAEARAE 692
	QY	359 LLDDAENETASILMSGFROMI-HMFNT-----ENPDQAAQOE-----395
	DB	693 ESM DAGIDAVEEVDAFTAVSDHADETDTGVQIEISDTTDDQAASTEAEVSMTEEVADLSD 752
	QY	396 -LAAQARAACAAGDSSAAAALADAQALEALGKAGOOQGIIN 437
	DB	753 STAGEAQSVSAAAEQ-AASMSEISDSVESLSGQAEQKALLS 794
	RESULT 5	
	HTR4_HALSA	STANDARD; PRT; 810 AA.
	ID	Q48317;
	DC	15-DEC-1998 (Rel. 37, Created)
	DT	20-AUG-2001 (Rel. 40, Last sequence update)
	DT	20-AUG-2001 (Rel. 40, Last annotation update)
	DT	HALOBACTERIAL TRANSDUCER PROTEIN IV.
	GN	HTR5 OR HTP1V.
	DE	Halobacterium salinarum.
	OS	Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
	OC	Halobacterium.
	OX	NCBI_Taxid=2242;
	RN	[1]
	RN	SEQUENCE FROM N.A.
	RP	STRAIN=S9;
	RC	
	RX	MEDLINE=96275896; PubMed=8674984;
	RA	Rudolph J., Nordmann B., Storch K.F., Gruenberg H., Rodewald K.,
	RA	Oesterheft D.;
	RL	"A family of halobacterial transducer proteins.";
	RT	FEMS Microbiol. Lett. 139:161-168(1996).


```

QY 59 SILRSV-----NALMSLADKLGIASSNSSSSSTSRSA-----DVSSTTATAPT 101
DB 396 RALESAADYEALTAADV--GDLTRVDSARHODAMARIGHALNDMLDDIETSVAAATA 453
QY 102 PPPPTSDYKTOAQATYDTTFTSTSLADIAQALVSLQDVAVTNKDPTAATDEETAIAEWE 161
DB 454 FSDHVSDAARVEADAGDAIDAGT---DVSTAVDEISDGATE-----QDRLHEVAGEVD 505
QY 162 TKNADAIKVGQAQITELAKYASDNQAIILDSLGKLTSPDLLQTLQSVANNKKAELLEM 221
DB 506 DLSASAEVAETVASLADTAGQAASAVDD--GROATEDAVET--MDDVADDAEAAADMDA 562
QY 222 QDNVPVP--GKTPAIAQSLVQDTATQIEKDGNAIGDAYFAGONASGAVENA-----273
DB 563 LDSEADIGEIVDVIADQTNMLAL-----NASTEAARTGADGDFAVVADEVKTLA 616
QY 274 -KSNNSISNIDSAKAATATAKTQIAEAKKPPSPILQEAQEMVIOAEKDLKNKPADGS 332
DB 617 EESRDAEEDIESRLALQGVSDVADEMRTSDT--VSDGRATVGDAAATLDDV-----668
QY 333 DVPNPGTTVGSKQGGSSIGSIRVSMILLDDAENETASILMSGFOMTHMENTENPDSQAA 392
DB 669 -----VSFVADTDFAAGQIRAA---TDQHAASRVASAVDEV-----AGI 706
QY 393 QOELAAQARA-AKAAGD-----DSAAAALADAQKALEAALGK 428
DB 707 SQETAQAQATAVDSAAATQDTLSSVDDAAADLADRAAALDDLLAE 751

RESULT 8
HTR2_HALSA
ID HTR2_HALSA STANDARD; PRT; 764 AA.
AC P71410;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONSES TO
DE SENSORY RHODOPSIN II TRANSDUCER (HTR-II) (METHYL-ACCEPTING PHOTOTAXIS
DE PROTEIN II) (MPP-II).
GN HTR2 OR HTR11.
OS Halobacterium salinarum.
OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
OC Halobacterium.
OX NCBI_TaxID=2242;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FLX15;
RX MEDLINE=96323203; PubMed=8710852;
RA Zhang W., Brooun A., Mueller M.M., Alam M.;
RA "The primary structures of the Archaeon Halobacterium salinarum blue
RT light receptor sensory rhodopsin II and its transducer, a methyl-
RT accepting protein.";
RL Proc. Natl. Acad. Sci. U.S.A. 93:8230-8235(1996).
CC -!- FUNCTION: TRANSDUCES SIGNALS FROM THE PHOTOTAXIS RECEPTOR
CC SENSORY RHODOPSIN II (SR-II) TO THE FLAGELLAR MOTOR. RESPONSES TO
CC LIGHT CHANGES THROUGH THE VARIATION OF THE LEVEL OF METHYLATION.
CC -!- SIMILARITY: BELONGS TO THE CHEMOTAXIS SENSORY TRANSDUCERS FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U62676; AAC4369.1;
DR InterPro; IPR000122; Chemotaxis_transducer.
DR InterPro; IPR000658; DUF5.
DR InterPro; IPR003660; HAMP.
DR Pfam; PF00672; DUF5; 1.
DR Pfam; PF00015; MCPsignal; 1.
DR SMART; SM00304; HAMP; 2.
DR SMART; SM00283; MA; 1.

```

```

KW Transducer; Photoreceptor; Transmembrane; Methylation.
FT INIT MET 0 0 BY SIMILARITY.
FT DOMAIN 1 15 CYTOPLASMIC (POTENTIAL).
FT TRANSMEM 16 36 POTENTIAL.
FT DOMAIN 37 277 EXTRACELLULAR (POTENTIAL).
FT TRANSMEM 278 298 POTENTIAL.
FT DOMAIN 299 764 CYTOPLASMIC (POTENTIAL).
SQ SEQUENCE 764 AA; 79187 MW; 1E0D7B4E60FC588 CRC64;

Query Match 6.4%; Score 154.5; DB 1; Length 764;
Best Local Similarity 22.6%; Pred. No. 0.44; Mismatches 93; Gaps 18;
Matches 105; Conservative 71;

QY 8 GPIDETERPPADLSAAGLEASAAKSAE-----AQLIAGABAKPE---SKTDSVERW 58
DB 337 GTLAESFRSRRDLSLSLTDARATARAEDAREDAEQORADAAAREDAEARKDAQEYA 396
QY 59 SILRSV-----NALMSLADKLGIASSNSSSSSTSRSA-----DVSSTTATAPT 101
DB 397 RALESAADYEALTAADV--GDLTRVDSARHODAMARIGHALNDMLDDIETSVAAATA 454
QY 102 PPPPTSDYKTOAQATYDTTFTSTSLADIAQALVSLQDVAVTNKDPTAATDEETAIAEWE 161
DB 455 FSDHVSDAARVEADAGDAIDAGT---DVSTAVDEISDGATE-----QDRLHEVAGEVD 506
QY 162 TKNADAIKVGQAQITELAKYASDNQAIILDSLGKLTSPDLLQTLQSVANNKKAELLEM 221
DB 507 DLSASAEVAETVASLADTAGQAASAVDD--GROATEDAVET--MDDVADDAEAAADMDA 563
QY 222 QDNVPVP--GKTPAIAQSLVQDTATQIEKDGNAIGDAYFAGONASGAVENA-----273
DB 564 LDSEADIGEIVDVIADQTNMLAL-----NASTEAARTGADGDFAVVADEVKTLA 617
QY 274 -KSNNSISNIDSAKAATATAKTQIAEAKKPPSPILQEAQEMVIOAEKDLKNKPADGS 332
DB 618 EESRDAEEDIESRLALQGVSDVADEMRTSDT--VSDGRATVGDAAATLDDV-----669
QY 333 DVPNPGTTVGSKQGGSSIGSIRVSMILLDDAENETASILMSGFOMTHMENTENPDSQAA 392
DB 670 -----VSFVADTDFAAGQIRAA---TDQHAASRVASAVDEV-----AGI 707
QY 393 QOELAAQARA-AKAAGD-----DSAAAALADAQKALEAALGK 428
DB 708 SQETAQAQATAVDSAAATQDTLSSVDDAAADLADRAAALDDLLAE 752

RESULT 9
SPAA_STRDO
ID SPAA_STRDO STANDARD; PRT; 1528 AA.
AC P21979;
DT 01-AUG-1991 (Rel. 19, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 01-FEB-1996 (Rel. 33, Last annotation update)
DE CELL SURFACE ANTIGEN I/II PRECURSOR.
GN SPAA.
OS Streptococcus downei (Streptococcus sobrinus).
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Streptococcus.
OX NCBI_TaxID=1317;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6715 / SEROTYPE G;
RX MEDLINE=91310320; PubMed=1855987;
RA Lapolla R.J., Haron J.A., Kelly C.G., Taylor W.R., Bohart C.,
RA Hendricks M., Pyati J., Graff R.T., Ma J.K.C., Lehner T.;
RA "Sequence and structural analysis of surface protein antigen I/II
RL (SpaA) of Streptococcus sobrinus.";
RL Infect. Immun. 59:2677-2685(1991).
RN [2]
RP SEQUENCE OF 423-817 FROM N.A.
RX MEDLINE=90299827; PubMed=1694526;
RA Goldschmidt R.M., Thoren-Gordon M., Curtiss R. III;

```


QY 48 KESTKDSVERWSILRSVAVNALMSLADKLGIASSN-SSSSTSRSDVSTTATATPTPPPT 106
DB 2 KVKTYGPRKSKISKTLCGAVLGTVAASVAGQKVFADETTTSDVTKVGTGTGNPAT 61
QY 107 -----SDDYKQATAYTFTST-----SLADIAQALVS-LQDVTNKT 147
DB 62 NLPEAQSGSAQKQSQTKLERQMVHTTEVPKTDLDQAADKAGSVNVQDADVN-KGT 120
QY 148 AATDEETAAEWE-----TKNADAIK-----VCAQITELAKYASDQATLDSIGK 193
DB 121 VKTAE-AYQKETEKTEDYQOAEIDKTTDOYKSDVAHAEVAKIKAKNQAFKEQYGG 179
QY 194 LTSFOLL-OTALLQSVANNKAAEL-----LKEMQDNPVPGKTPAIAQSLVDQTDAT 247
DB 180 -----DWVAKHAEVERINNAASKTAYEAKLAQYQADLAQVOKTNAANQASY-QKALAA 234
QY 248 QTE-----KGNAGIDAYFAGNAGAVENAKSNNSISNIDSAKAIATAK-----TQTAEA 299
DB 235 QALKRQVEANAAKAAAY---DTAVANNAKNTETAANEEIRKRNATAKAEYETKLAQY 291
QY 300 Q---KKFPDSPILOAE-OMVIOA-EKDLKNIKPAD-----GSDVPNPGTTVG 342
DB 292 QALKRQVEANAAEADYQAKLTAYQTELARVORANADAKAAYEAANNAKNAALTAE 351
QY 343 GS-----KQGGSIGSIRVSMILLDDAENETASILMSGFPMHMFNTE 385
DB 352 NTAIKORNENAKATYEAALKOYEADLAAVKANAANEADYQAK---LTAYQTELARVOKA 408
QY 386 NPDQAAQOELAAQARAQAA--GDSAAAL-ADAQKALEAALGK 428
DB 409 NADAKAAYEAANAANAALTAENTAIAKRNADAKADYEAALAK 454

RESULT 11
FLJB_SALTY STANDARD; PRT; 505 AA.
AC P52616; P97159;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PHASE-2 FLAGELLIN.
GN FLJB OR H2.
OS Salmonella typhimurium.
OC Bacteria; proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-SL 375;
RX MEDLINE=95325331; PubMed=7541401;
RA Vanegas R.A., Joys T.M.;
RT "Molecular analyses of the phase-2 antigen complex 1,2,... of
Salmonella spp.",
RL J. Bacteriol. 177:3863-3864(1995).
RN [2]
RP SEQUENCE OF 1-37 FROM N.A.
RX MEDLINE=82049491; PubMed=6271461;
RA Silverman M., Zieg J., Mandel G., Simon M.;
RT "Analysis of the functional components of the phase variation
system",
RL Cold Spring Harb. Symp. Quant. Biol. 45:17-26(1981).
RN [3]
RP SEQUENCE OF 482-505 FROM N.A.
RC STRAIN-SJ2353;
RA Minorance J., Tanaka S., Tomimaga A., Enomoto M.;
RL Submitted (JAN-1994) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: FLAGELLIN IS THE SUBUNIT PROTEIN WHICH POLYMERIZES TO
FORM THE FILAMENTS OF BACTERIAL FLAGELLA.
CC -!- MISCELLANEOUS: INDIVIDUAL SALMONELLA SEROTYPES USUALLY ALTERNATE
BETWEEN THE PRODUCTION OF 2 ANTIGENIC FORMS OF FLAGELLIN, TERMED
PHASE-1 AND PHASE-2, EACH SPECIFIED BY SEPARATE STRUCTURAL GENES.
CC -!- SIMILARITY: BELONGS TO THE BACTERIAL FLAGELLIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
or send an email to license@isb-sib.ch).

DR EMBL: U17177; AAC43354.1; -
DR EMBL: V01370; CAA24655.1; -
DR EMBL: D26168; BAA05156.1; -
DR StyGene; SG10564; FLJB.
DR InterPro; IPR001029; Flagellin_C.
DR InterPro; IPR001492; Flagellin_N.
DR Pfam; PF00700; Flagellin_C; 1.
DR Pfam; PF00669; Flagellin_N; 1.
DR PRINTS; PR00207; FLAGELLIN.
DR ProDom; PD000316; Flagellin_C; 1.
DR Flagella.
KW INIT_MET 0 0 BY SIMILARITY.
FT CONFLICT 37 37 I -> S (IN REF. 2).
SQ SEQUENCE 505 AA; 52404 MW; 3A5CC404AF7AF88B CRC64;

Query Match: 6.1%; Score 148; DB 1; Length 505;
Best Local Similarity 21.0%; Pred. No. 0.56;
Matches 93; Conservative 59; Mismatches 168; Indels 122; Gaps 20;

QY 23 AQGLEASANKSABEAQRIAG-----AEAPKESKTDSEVERWSILRSVAVNALMSLADKL 75
DB 32 SSGLRINSKADDAAGQAIAANFTANIKGLTQASRNANDGISIAQTTEGALNEINNLRV 91
QY 76 GIASSNSSSTSRSDVSTTATATPTPPPTSDDYKTAQ-----TAYDTFT----- 123
DB 92 RELAVQSANSTNSQSLDSIOAEI-TQRLNEIDRVSGTQFNGVKVLAQDNTLTIVGAN 150
QY 124 -----STSLADIQAALVSLQ-----DAVTNINKDTAATD-----EETAAE 159
DB 151 DGETIDIDLKQINSQTILGSLNVQKAYDKVDTAVTTKAYANNGTTLDSGLDDAAIKAA 210
QY 160 WETKNADAIVGAOITELAKYASDQALDSIGKLTSLFDLLQ-----TALLQS 207
DB 211 TGGTNGTASVTGGAV-----KFDADNNKYFTVIGGTGADAAGKNGDYEVNATDGTVLA 266
QY 208 VANNN-----KAAELKEMQDNPV---PGKTPAIAQSLVDQTDATATQI-----EKD 252
DB 267 GATKTPAGATTTEVQELKDTPAVVSADAKNALIAGG-VDATDANGAELVKMSYTDKN 325
QY 253 GNAL-----GDAYFAG--QNASGAVENAKSN-----NSISNIDSAKAAI 289
DB 326 KKTIEGGYALKRAGDKYYAADYDEATGAIK-AKTTSTYTAADGTTKTAANQLGGVD-GKTEV 383
QY 290 AT-----AKTQIAEAKQKFPDPSILQEAQEMVITQAE---KDLKNIKPA 329
DB 384 VTIDGKTYNASKAAGHDFKAQPELAEEAAKTENP-LOKIDAALAAQVDALESLDGLGAVNR 442
QY 330 GDSVPNPGTTVGSKQGGSSI 351
DB 443 FNSAITNLGNTVNNLSEARSRI 454

RESULT 12
YJH8_YEAST STANDARD; PRT; 881 AA.
AC P47033;
DT 01-FEB-1996 (Rel. 33, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 89.2 KDA PROTEIN IN SCPL160-SMC3 INTERGENIC REGION.
GN YJL078C OR J1027.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.

QY 238 LVDTDTATATQIEKDN--AIGDAY--FAGQNASGAVE-----NAKSNNSISNIDS 284
 Db 603 LALNANTEAARADQDGGFAVAVADEVKDLADESKASAAEIALVAEVRAGTQETSVAMDR 662
 QY 285 AKAAIATAKTAQIAEAOQKFFDPSILOEAEQWIOAE---KDLKNIRKPADGSDVPNPGTTV 341
 Db 663 IOERVSGVETVSTERS-----LSEIAGRIAEADTGVGEISNAMDDQASVSDVTTAV 716
 QY 342 GGSQKQSSGSSIRVSMMLDDAENETASILMSGFROMIHMFTEN-PDSQAAQOEIAAQA 400
 Db 717 GDVAALGEETATEASTA--DAAAEQASTLSDVAAQ-----TETLAHAVALREHAAQF 768
 QY 401 RAA 403
 Db 769 EVA 771

RESULT 14

STF_LAMBD STANDARD; PRT; 774 AA.
 AC P03764; P03745;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 20-AUG-2001 (Rel. 40, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE SIDE TAIL FIBER PROTEIN.
 GN STF.
 OS Bacteriophage lambda.
 OC Viruses; dsDNA viruses, no RNA stage; Tailed phages; Siphoviridae;
 OC Lambda phage group.
 OX NCBI_TaxID=10710;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83189071; PubMed=62211115;
 RA Sanger F., Coulson A.R., Hong G.F., Hill D.F., Petersen G.B.;
 RT "Nucleotide sequence of bacteriophage lambda DNA.";
 RL J. Mol. Biol. 162:729-773(1982).
 RN [2]
 RP IDENTIFICATION AS STF.
 RX MEDLINE=92165720; PubMed=1531648;
 RA Haggaard-Ljungquist E., Hailing C., Calendar R.;
 RT "DNA sequences of the tail fiber genes of bacteriophage P2: evidence for horizontal transfer of tail fiber genes among unrelated bacteriophages.";
 RL J. Bacteriol. 174:1462-1477(1992).
 RN [3]
 RP RECONSTRUCTION OF STF.
 RX MEDLINE=93068310; PubMed=1439823;
 RA Hendrix R.W., Duda R.L.;
 RT "Bacteriophage lambda P2a: not the mother of all lambda phages.";
 RL Science 258:1145-1148(1992).
 CC -!- MISCELLANEOUS: THE COMMON LABORATORY STRAIN OF BACTERIOPHAGE LAMBDA; LAMBDA P2a; CARRIES A FRAMESHIFT MUTATION RELATIVE TO UR-LAMBDA, THE ORIGINAL ISOLATE. THE UR-LAMBDA VIRIONS HAVE THIN, JOINTED TAIL FIBERS (SIDE TAIL FIBERS) THAT ARE ABSENT FROM LAMBDA WILD TYPE. RELATIVE TO LAMBDA P2a, UR-LAMBDA HAS EXPANDED RECEPTOR SPECIFICITY AND ADSORBS TO E. COLI CELLS MORE RAPIDLY.
 CC -!- SIMILARITY: BELONGS TO THE TAIL FIBER FAMILY.
 CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS CORRECTED IN POSITION TO 396 TO RECREATE THE ORIGINAL STF PROTEIN.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license@isb-sib.ch. See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch.
 CC -----
 CC EMBL: J02459; AAA96555.1; ALT_FRAME.
 DR EMBL: J02459; AAA96557.1; ALT_FRAME.
 DR PIR: A04389; QXBP1L.
 DR PIR: A04370; QXBP2L.

DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR001778; POA_allergen.
 KW Fiber protein.
 SQ SEQUENCE 774 AA: 77527 MW: CDD1DF85B919123B CRC64;

Query Match 6.1%; Score 146.5; DB 1; Length 774;
 Best Local Similarity 20.9%; Pred. No. 1.1;
 Matches 90; Conservative 66; Mismatches 209; Indels 65; Gaps 13;
 QY 26 LEASAANKSAEAOIAGAEAKPKESKTSVERVSIILRSVAVNALMSLADKLGITASSNSSS 85
 Db 110 VEEVARNASVVGQSTADAKKSGADASAAQVAALVTDATDSARAASTSAGQAAAS-SAGE 168
 QY 86 TERSADVSTTATPTPPPTSDDYKTOAQATVDTFTTSTSLADIQAALVSLQDAVNIK 145
 Db 169 ASSGAEEASAKATEAEKSAAAESSKNAATSAAGAAKTSET-----NAAASQSAATSAS 223
 QY 146 DTAATDEETAIAAEWETKNADAIVGAQITELAKYASDQAILDSLGLKLTSLFDLLQALL 205
 Db 224 TAATKASEAATSA---RDVAASKEAAKSSSETNASSAGRAASSATAAENSARAAKTS-- 277
 QY 206 QSVANNKKAELLKEMQNPVPGKTPAIAQSLVDOTDATATQIEKDGNAIGDAYFAGQN 265
 Db 278 ---ETNARSSETAERSASAAADAKTAAGSASTATKAT-----EAAGSAVSASOS 326
 QY 266 ASGAVENA-KSNNSISNIDSAAKAAIA-----TAKTOIAEAAQKKFPDPSILQEAQWVIOA 319
 Db 327 KSAEAAAIRAKSKRAEDIASAVALEADTTKRGIVQLSSATNSITSETLAATPAKVAV 386
 QY 320 EKDLKNIK-PADG---SDVNPNGTIVGGSKQKQSSIGSIRVSMMLDDAENETASILMSGF 375
 Db 387 VMDETNRKAPLDSPALVTGTPTAPTALRGTT-----NNTQIANTAFVLAAL 430
 QY 376 RQMIHMFTENPDQSAQAQELAAQAAKAAAGDDSAALADAAQKALEAALGALGAGQOOGI 435
 Db 431 ADVI-----DASPDALNTLNELAA-----ALGND-----PDFATTMTNALAQKPRNAT 474
 QY 436 LNALGOIASA 445
 Db 475 LTALAGLSTA 484

RESULT 15

SYEP_DROME STANDARD; PRT; 1714 AA.
 AC P28668; Q9VCF5;
 DT 01-DEC-1992 (Rel. 24, Created)
 DT 01-DEC-1992 (Rel. 24, Last sequence update)
 DT 20-AUG-2001 (Rel. 40, Last annotation update)
 DE BIFUNCTIONAL AMINOACYL-TRNA SYNTHETASE [INCLUDES: GLUTAMYL-TRNA DE SYNTHETASE (EC 6.1.1.17) (GLUTAMATE--TRNA LIGASE); PROLYL-TRNA DE SYNTHETASE (EC 6.1.1.15) (PROLINE--TRNA LIGASE)].
 GN AATS-GLUPRO OR CG5394.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92097547; PubMed=1756734;
 RA Cerini C., Kerjan P., Astier M., Gratecos D., Mirande M., Semeriva M.;
 RT "A component of the multisynthetase complex is a multifunctional aminoacyl-tRNA synthetase.";
 RL EMBL J. 10:4267-4277(1991).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=OREGON-R;
 RX MEDLINE=97217441; PubMed=9063462;
 RA Cerini C., Semeriva M., Gratecos D.;
 RT "Evolution of the aminoacyl-tRNA synthetase family and the organization of the Drosophila glutamyl-prolyl-tRNA synthetase gene.

RT Intron/exon structure of the gene, control of expression of the two
 RL mRNAs, selective advantage of the multienzyme complex.;
 Eur. J. Biochem. 244:176-185(1997).

RP [3].
 RN SEQUENCE FROM N.A.

RC STRAIN=BERKELEY;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Randell M.D., Zhang Q., Chen L.A.,
 RA Brandon R.C., Rogers J.-H.C., Blazie R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abrial J.F., Agbayani A., An H.-J., Andrews-Pfannkuch C., Baldwin D.,
 RA Ballew R.M., Basu A., Bakendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borkova D., Botchan M.R., Bouck J., Brokstein P., Brotier P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahike C., Davenport L.B., Davies P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
 RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Fosler C., Gabriellian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
 RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacieb J.M.,
 RA Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
 RA Shue B.C., Sidenkiamos I., Simpson M., Skupski M.P., Smith T.,
 RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
 RA Swirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
 RA Wang Z.-Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster";
 RL Science 287:2185-2195(2000).
 CC -1- CATALYTIC ACTIVITY: ATP + L-GLUTAMATE + TRNA(GLU) = AMP +
 CC PYROPHOSPHATE + L-GLUTAMYL-TRNA(GLU).
 CC -1- CATALYTIC ACTIVITY: ATP + L-PROLINE + TRNA(PRO) = AMP +
 CC PYROPHOSPHATE + L-PROLYL-TRNA(PRO).
 CC -1- SUBUNIT: COMPONENT OF THE MULTISYNTHETASE COMPLEX WHICH IS
 CC COMPRISED OF A BIFUNCTIONAL GLUTAMYL-PROLYL-TRNA SYNTHETASE, THE
 CC MONOSPECIFIC ISOLEUCYL, LEUCYL, GLUTAMINYL, METHIONYL, LYSYL,
 CC ARGINYL, AND ASPARTYL-TRNA SYNTHETASES AS WELL AS THREE AUXILIARY
 CC PROTEINS, P18, P48 AND P43.
 CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO CLASS-I
 CC AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO CLASS-II
 CC AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -1- SIMILARITY: CONTAINS 6 "WHEP-TRS" DOMAINS.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL; M74104; AAC28594.1;
 CC DR EMBL; U59923; AAC47469.1;
 CC DR EMBL; AE003745; AAF56211.1;
 CC DR PIR; S18644; S18644.
 CC HSSP; P00962; IQRU.

DR FlyBase; FBgn0005674; Aats-glupro.
 DR InterPro; IPR002106; AA_TRNA_ligase_II.
 DR InterPro; IPR000738; WHEP-TRS.
 DR InterPro; IPR000924; tRNA-synt_1c.
 DR InterPro; IPR002314; tRNA-synt_2b.
 DR InterPro; IPR001412; tRNA-synt_1.
 DR InterPro; IPR002316; tRNA-synt_pro.
 DR Pfam; PF00749; tRNA-synt_1c; 1.
 DR Pfam; PF00587; tRNA-synt_2b; 1.
 DR Pfam; PF00458; WHEP-TRS; 6.
 DR PRINTS; PR00987; TRNASYNTHGLU.
 DR PRINTS; PR01046; TRNASYNTHPRO.
 DR PROSITE; PS00178; AA_TRNA_LIGASE_I; 1.
 DR PROSITE; PS00179; AA_TRNA_LIGASE_II_1; FALSE_NEG.
 DR PROSITE; PS00339; AA_TRNA_LIGASE_II_2; FALSE_NEG.
 DR PROSITE; PS00762; WHEP-TRS; 6.
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Multifunctional enzyme; Repeat.
 FT DOMAIN 170 754 GLUTAMYL-TRNA SYNTHETASE.
 FT DOMAIN 755 800 WHEP-TRS 1.
 FT DOMAIN 827 872 WHEP-TRS 2.
 FT DOMAIN 901 945 WHEP-TRS 3.
 FT DOMAIN 980 1025 WHEP-TRS 4.
 FT DOMAIN 1055 1100 WHEP-TRS 5.
 FT DOMAIN 1129 1173 WHEP-TRS 6.
 FT DOMAIN 1174 1180 POLY-GLY.
 FT DOMAIN 1207 1714 PROLYL-TRNA SYNTHETASE.
 FT SITE 209 220 "HIGH" REGION.
 FT SITE 438 442 "RMSKS" REGION.
 FT BINDING 441 441 ATP (BY SIMILARITY).
 FT CONFLICT 102 106 TSPLP -> DKSTIA (IN REF. 3).
 FT CONFLICT 233 234 VC -> AF (IN REF. 3).
 FT CONFLICT 341 345 NTACA -> KYCVR (IN REF. 3).
 FT CONFLICT 583 583 K -> R (IN REF. 3).
 FT CONFLICT 692 692 L -> A (IN REF. 3).
 FT CONFLICT 753 753 T -> S (IN REF. 3).
 FT CONFLICT 802 802 T -> S (IN REF. 3).
 FT CONFLICT 873 873 P -> T (IN REF. 3).
 FT CONFLICT 887 887 G -> V (IN REF. 3).
 FT CONFLICT 1201 1201 P -> PA (IN REF. 3).
 FT CONFLICT 1461 1461 MISSING (IN REF. 3).
 FT CONFLICT 1587 1587 G -> V (IN REF. 3).
 SQ SEQUENCE 1714 AA; 189197 MW; 6FE8C58045E48AC CRC64;

Query Match 6.1%; Score 146; DB 1; Length 1714;
 Best Local Similarity 19.8%; Pred.No. 3.2; Indels 142; Gaps 20;
 Matches 101; Conservative 74; Mismatches 192;

QY 7 PGFI-----DETERTPPADLSAQGLEASAAANKSAAQAQRIAG-----A 43
 DB 701 PSFVLVFSIPDGHGTHKDVPTSGLKVNAPDAKATKASSPVSSGQASELDSQITQGGDLVR 760
 QY 44 EAPKESKTSVSWERWTLRSVAVNALMSLADKLGIASSNSSSTSRSDVSTTAT-APTP 102
 DB 761 DLAKSKAAKQDID-----VAVKKLLAL-----KADYKSATGKWPGQTSAPVP 807
 QY 103 PPPTSDDYKTOAOATYDTFTSTSLADIQAALVSLODAVNIKDTAATDEATA----- 157
 DB 808 AASSS-----SANDAV-----SVNASIVKQGGDLVRDLKGGKASKPEIDAIVKTL 852
 QY 158 ---AEWETKNADAIVKGAQITELAKYASDNQAI--LDSLGLKLSFDLLQTLALQSVAANN 212
 DB 853 ELKAAQYKTLTGQDWKPKCTVPTTAAPSASAPSVGVGNSVQAILS-----QITAGD 903
 QY 213 KAELLKEMODNPVVPKCTPTAIAQSLVDQDTATQTEKGNALGDYFACONASGAVEN 272
 DB 904 KVRRLKSAKADKATVDAAVAVKTLVLSLKADYKAATGSD-WKPGTT-----APAPAAAPVK 956
 QY 273 AKSN-----SISNIDSAKAIAIATAKTAIAEAKKFPDPSPILOEAEQMVIOAEKDKNI-- 326
 DB 957 KQKNPDPASVLIVNTLLNKIAQGGDKIRQLKSAKSEKSLVEAEVKULLALKTDYKSLTG 1016

QY 327 ---KPADGSDVPNPGTT-----VGSKQOGSSIGSI----- 354
Db 1017 QEWKP--GTVPAPTIVNVIDLTTGG--DGSVDYGSVLKIQAGDKIRKLKSEKAAKNVI 1072
QY 355 --RVSMLL-----DDAENETASILMSGFROMIHMFNTENPDSQAAQOELA 397
Db 1073 DPEVKTLALKGEYKTLGSKDWTPTDAKSEPAVV-----KKEASPVSMASPAKDELTOEIN 1127
QY 398 AQARAAKAGDSDSAAALADAQALEAAL 426
Db 1128 AQGEKVRAGKNGKAKEVIDAEVAKLLAL 1156

Search completed: February 7, 2002, 21:42:42
Job time: 596 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:41:09 ; Search time 172 Seconds
(without alignments)
416.706 Million cell updates/sec

Title: US-09-391-606-8
Perfect score: 2412
Sequence: 1 MVNPIGPGIDETERTPPAD.....QKLISEEDLSAVDHHHHH 490

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 473505 seqs, 146272329 residues
Total number of hits satisfying chosen parameters: 473505

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPREMBL_17.*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mmc.*
- 8: sp_organelle.*
- 9: sp_phase.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2214	91.8	715	2 Q46166	Q46166 chlamydia p
2	2188	90.7	651	2 Q927H7	Q927H7 chlamydia p
3	451.5	18.7	647	2 Q84627	Q84627 chlamydia t
4	195.5	8.1	1327	2 Q9X7M2	Q9X7M2 staphylococ
5	195.5	8.1	2478	2 Q9RL69	Q9RL69 staphylococ
6	193.5	8.0	2478	2 Q9LCH2	Q9LCH2 staphylococ
7	188	7.8	2055	2 Q85472	Q85472 abirotrophia
8	182.5	7.6	2481	2 Q99QR6	Q99QR6 staphylococ
9	182	7.5	1545	2 Q9RDO1	Q9RDO1 streptomyce
10	181	7.5	697	5 Q9NDJ0	Q9NDJ0 plasmodium
11	180	7.5	2271	2 Q99QY4	Q99QY4 staphylococ
12	178	7.4	1822	2 Q07290	Q07290 streptococ
13	174.5	7.2	1795	2 Q9LCJ9	Q9LCJ9 staphylococ
14	173	7.2	2016	5 Q9BIT0	Q9BIT0 plectureys
15	171	7.1	627	1 Q9HS86	Q9HS86 halobacteri
16	171	7.1	993	2 Q99QZ5	Q99QZ5 staphylococ
17	170	7.0	1128	5 Q26947	Q26947 trypanosoma
18	168.5	7.0	1038	10 Q9AS09	Q9AS09 oryza sativ
19	166.5	6.9	2232	5 P91365	P91365 caenorhabdi

20	166.5	6.9	6713	2 Q99U54	Q99U54 staphylococ
21	163.5	6.8	956	2 Q9LON7	Q9LON7 streptomyce
22	163.5	6.8	2045	2 Q9AOK5	Q9AOK5 streptococ
23	161.5	6.7	3381	2 Q9KX33	Q9KX33 streptococ
24	160.5	6.7	344	1 P71409	P71409 halobacteri
25	160.5	6.7	1156	2 Q925A4	Q925A4 streptomyce
26	160.5	6.7	2178	2 Q9KWR3	Q9KWR3 streptococ
27	159	6.6	1579	11 Q99WP1	Q99WP1 mus musculu
28	159	6.6	1684	11 Q9WTQ5	Q9WTQ5 mus musculu
29	158	6.6	1365	2 Q49525	Q49525 mycoplasma
30	158	6.6	2186	2 Q99TB0	Q99TB0 staphylococ
31	157.5	6.5	1029	10 Q9LDB1	Q9LDB1 oryza sativ
32	157	6.5	1344	2 Q49545	Q49545 mycoplasma
33	157	6.5	6677	5 Q9N435	Q9N435 caenorhabdi
34	156.5	6.5	641	2 Q9KS57	Q9KS57 vibrio chol
35	156	6.5	845	5 Q9Y1P8	Q9Y1P8 plasmodium
36	156	6.5	990	13 Q91803	Q91803 xenopus lae
37	156	6.5	993	10 Q9ASL3	Q9ASL3 oryza sativ
38	155.5	6.4	582	5 Q16967	Q16967 aplysia cal
39	155.5	6.4	1041	10 Q9ASA4	Q9ASA4 oryza sativ
40	155.5	6.4	2273	5 Q9U141	Q9U141 leishmania
41	155	6.4	892	3 Q9P3P5	Q9P3P5 neospora
42	155	6.4	969	5 Q9ND19	Q9ND19 plasmodium
43	155	6.4	1147	2 Q87848	Q87848 streptomyce
44	155	6.4	1569	2 Q54183	Q54183 streptococ
45	154.5	6.4	1262	5 Q20684	Q20684 caenorhabdi

ALIGNMENTS

RESULT 1					
Q46166	Q46166	PRELIMINARY;	PRT;	715 AA.	
AC	Q46166;				
DT	01-NOV-1996	(TRENBLrel. 01, Created)			
DT	01-NOV-1996	(TRENBLrel. 01, Last sequence update)			
DT	01-MAY-2000	(TRENBLrel. 13, Last annotation update)			
DE	76 KDA PROTEIN.				
OS	Chlamydia pneumoniae	(Chlamydia pneumoniae)			
OC	Bacteria; Chlamydiales; Chlamydiaceae; Chlamydia				
OX	NCBI_TaxID=83558;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RA	MEDLINE=94156481; PubMed=75093320;				
RX	Perez-Melgosa M., Kuo C.C., Campbell L.				
RT	*Isolation and characterization of a gene encoding a Chlamydia				
RT	pneumoniae 76-kilodalton protein containing a species-specific				
RT	epitope.				
RL	Infect. Immun. 62:880-886(1994).				
DR	EMBL; L23921; AAA23117.1;				
SQ	SEQUENCE 715 AA; 76626 MW; 3F01C9A59DE964C5 CRC64;				

Query Match	91.8%;	Score 2214;	DB 2;	Length 715;	
Best Local Similarity	98.9%;	Pred. No. 1.8e-114;			
Matches	454;	Conservative	2;	Mismatches	3;
				Indels	0;
				Gaps	0;
QY	1	MVNPIGPGIDETERTPPADLSAOGLEASAAANKSAEORIAAGAKPKESKTSVERWSI	60		
Db	257	LVNPIGPGIDETERTPPADLSAOGLEASAAANKSAEORIAAGAKPKESKTSVERWSI	316		
QY	61	LRSAVNALMSLADKLGIASSNSSSTSRKSDVDSTTATATPTPPPTSDYKTAQTAYDT	120		
Db	317	LRSAVNALMSLADKLGIASSNSSSTSRKSDVDSTTATATPTPPPTSDYKTAQTAYDT	376		
QY	121	IFTSTSLADIQAALVLSQDAVTNFKDAAETAEAEWETKNADAIKVAQITELAKY	180		
Db	377	IFTSTSLADIQAALVLSQDAVTNFKDAAETAEAEWETKNADAIKVAQITELAKY	436		
QY	181	ASDNOAILDLSGLKLTSTFDLLQALLOQVANNKKAELLKEMQDNVPVPGKTPATAQSLVD	240		
Db	437	ASDNOAILDLSGLKLTSTFDLLQALLOQVANNKKAELLKEMQDNVPVPGKTPATAQSLVD	496		

QY 241 QTDATATQIEKDGNAICDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 300
 Db 497 QTDATATQIEKDGNAICDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 556
 QY 301 KKFPPSPILQEAQOMVIOAEKDLKNKIPADGSDVNPNGTGVGSKQOQSSIGSIRVSMLL 360
 Db 557 KKFPPSPILQEAQOMVIOAEKDLKNKIPADGSDVNPNGTGVGSKQOQSSIGSIRVSMLL 616
 QY 361 DDAENETASILMSGFQMIHMFNTENPDSQAQOELAAQARAARAKAGDDSAALADAK 420
 Db 617 DDAENETASILMSGFQMIHMFNTENPDSQAQOELAAQARAARAKAGDDSAALADAK 676
 QY 421 ALAALGKAGQOQGIILNALGQIASAAVVSAGVLPLOQVL 459
 Db 677 ALAALGKAGQOQGIILNALGQIASAAVVSAGVLPLOQVL 715

RESULT 2

Q9Z7H7 PRELIMINARY; PRT; 651 AA.
 AC Q9Z7H7;
 DT 01-MAY-1999 (TrEMBLrel. 10, Created)
 DT 01-MAY-1999 (TrEMBLrel. 10, Last sequence update)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
 DE CHLPN 76 KDA HOMOLOG_1 (CT622).
 GN CPN0728 OR CPJ0728 OR CP0018.
 OS Chlamydia pneumoniae (Chlamydia pneumoniae).
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.
 OX NCBI_TaxID=83558;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=J138;
 RX MEDLINE=20330349; PubMed=10192388;
 RA Kalman S., Mitchell W., Marathe R., Lammel C., Fan J., Hyman R.W.,
 RA Shiba T., Ishii K., Hattori M., Kuhara S., Nakazawa T.;
 RA "Comparative genomics of whole genome sequences of Chlamydia pneumoniae J138
 from Japan and CWL029 from USA".
 RL Nucleic Acids Res. 28:2311-2314(2000).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=AR39;
 RX MEDLINE=20150255; PubMed=10684935;
 RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,
 RA White O., Hickey E.K., Peterson J., Umayam L.A., Utterback T.,
 RA Berry K., Bass S., Linher K., Weidman J., Khouri H., Craven B.,
 RA Bowman C., Dodson R., Gwin M., Nelson W., DeBoy R., Kolonay J.,
 RA McClarty G., Salzberg S.L., Eisen J., Fraser C.M.;
 RA "Genome sequences of Chlamydia trachomatis MoPn and Chlamydia
 pneumoniae AR39".
 RL Nucleic Acids Res. 28:1397-1406(2000).
 DR EMBL; AE001654; AAD18867.1;
 DR EMBL; AF002547; BAA98935.1;
 DR EMBL; AE002165; AAF37914.1;
 DR TIGR; CP0018;
 KW Complete proteome.
 SQ SEQUENCE 651 AA; 68217 MW; 47AE6C3FF2FF0123 CRC64;

Query Match 90.7%; Score 2188; DB 2; Length 651;
 Best Local Similarity 98.9%; Pred. No. 4.4e-113;
 Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MWNPIGPGIDERTPPADLSAQGLEASAAKSAQRIAGAEAKPKESKTDTSVERWSI 60
 |||||

Db 1 MWNPIGPGIDERTPPADLSAQGLEASAAKSAQRIAGAEAKPKESKTDTSVERWSI 60
 QY 61 LRSAYNALMSLADKLGIASSNSSSTSRSDVDTTATPTPPPTSDYKTQATAYDT 120
 Db 61 LRSAYNALMSLADKLGIASSNSSSTSRSDVDTTATPTPPPTSDYKTQATAYDT 120
 QY 121 IFTSTSLADIOAALVSLQDAVNIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 180
 Db 121 IFTSTSLADIOAALVSLQDAVNIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 180
 QY 181 ASDNQAILDSLGLTSTFDLLQALLOSVANNKKAELLKEMODNPVPGKPAIAQSLVD 240
 Db 181 ASDNQAILDSLGLTSTFDLLQALLOSVANNKKAELLKEMODNPVPGKPAIAQSLVD 240
 QY 241 QTDATATQIEKDGNAICDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 300
 Db 241 QTDATATQIEKDGNAICDAYFAGQNASGAVENAKSNNSISNIDSAAKATATAKTQIAEAQ 300
 QY 301 KKFPPSPILQEAQOMVIOAEKDLKNKIPADGSDVNPNGTGVGSKQOQSSIGSIRVSMLL 360
 Db 301 KKFPPSPILQEAQOMVIOAEKDLKNKIPADGSDVNPNGTGVGSKQOQSSIGSIRVSMLL 360
 QY 361 DDAENETASILMSGFQMIHMFNTENPDSQAQOELAAQARAARAKAGDDSAALADAK 420
 Db 361 DDAENETASILMSGFQMIHMFNTENPDSQAQOELAAQARAARAKAGDDSAALADAK 420
 QY 421 ALAALGKAGQOQGIILNALGQIASAAVVSAGVLP 454
 Db 421 ALAALGKAGQOQGIILNALGQIASAAVVSAGVLP 454
 RESULT 3
 O84627 PRELIMINARY; PRT; 647 AA.
 ID O84627;
 AC O84627;
 DT 01-NOV-1998 (TrEMBLrel. 08, Created)
 DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
 DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)
 DE CHLPN 76 KDA HOMOLOG.
 GN CT622.
 OS Chlamydia trachomatis.
 OC Bacteria; Chlamydiales; Chlamydiales; Chlamydiales.
 OX NCBI_TaxID=813;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=D/UV-3/CX;
 RX MEDLINE=99000809; PubMed=97841136;
 RA Stephens R.S., Kalman S., Lammel C.J., Fan J., Marathe R., Aravind L.,
 RA Mitchell W.P., Olinger L., Tatusov R.L., Zhao Q., Koonin E.V.,
 RA Davis R.W.;
 RA "Genome sequence of an obligate intracellular pathogen of humans:
 Chlamydia trachomatis".
 RL Science 282:754-759(1998).
 DR EMBL; AE001333; AAC68226.1;
 KW Complete proteome.
 SQ SEQUENCE 647 AA; 68525 MW; C0D14C2D74473625 CRC64;

Query Match 18.7%; Score 451.5; DB 2; Length 647;
 Best Local Similarity 29.6%; Pred. No. 2.3e-17;
 Matches 138; Conservative 85; Mismatches 192; Indels 51; Gaps 14;
 QY 2 VNPPIGPGIDERTPPADLSAQGLEASAAKSAQRIAGAEAKPKESKTDTSVERWSI 61
 |||||
 Db 15 MNPPIGQI-----ASNSETKESTKESEA-----SPSASSSSWSFL 52
 QY 62 RSAYNALMSLADKLGIASSNSSSTSRSDVDTTATPTPPPTSDY---KTQATAY 118
 Db 53 SSAAHALISLRD--AILNKNSSPTDSLS--OLEASTSIS--TVTRVAARDYNEAKSNFTAK 108
 QY 119 DTFTSTSLADIOAALVSLQDAVNIKDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 178
 Db 109 SGLNATTLAETKMKADLMAALQDMERLAKQKAEVTRIKEALQOEKQEVID---KLNLV 165


```
Db 838 KMDAYNEVKQAATARKQATVSNATNEEVAEADAADVAQAQKGLHDIQVVKSKQEVADT 897
Qy 88 RSADVDSSTA-TAPTPPPPTSDYKYQAQPAYDT-----IFTSTSLADIQAALVSL--- 137
Db 898 KSKVLDRKINAIOQAKVKAAD---TEVENAYNTRKQEIQNSNASTTEKQAAAYTELDTK 954
Qy 138 QDAYTNIKDTAATD-----EETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDS 190
Db 955 KQEARTNL-DAANTNSDVTTAKDNTSIAAINQVQAATTKKSDAK-AEIAQKASERKTAIEA 1012
Qy 191 LGKLT-----SFDLIQTALLOSVAANNKAAELLKEMQDNPPVPGKT----- 231
Db 1013 MNDSTTEEQAAKDKVDQAVVTANADIDNAA-----ANNVDNAKTNTNEATIAATTPDA 1066
Qy 232 ---PATAQSLVDQDTATATQIEKDN-----AIGDAYFAGQNASGA 269
Db 1067 NVKPAAKQAIAQKQVQAAQETAI--DGNNGSTTEEKAQAQVQTEKTITADAAIDAHAHTNAE 1124
Qy 270 VENAKSNNSISNI-----DSAKAAIAT---AKTOIAEAOAKFFPDSPILOEA--E 313
Db 1125 VEAAK-KAAIAKIEAIOQATTTKDNKAKEIATKANERKTAIAQODITAEIAAANDVD 1183
Qy 314 QMWIOAEKDLKNIKPADG-SDVPNPGTTVGGSKQGGSSIGSIRVSMLLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVDQAKTTGENSIDQVTPVNNKAT-----ARNEITAILN 1235
Qy 373 SGFRQMIHMFNTENPDSQAQOELAAQARAAGAAGDSAAA-----ALADAQAQALEAA 425
Db 1236 NKLEIQATPDATDEEKQAADE--ANTENGKANQAISAATTNAQVDEKANAEEAIAINAV 1293
Qy 426 LGKAGQOQGLNALGOITASAAVVSAGVLPLOQVLRIRARYQAVVEOKLISEEDLNSAV 483
Db 1294 TPKVVVKQAQKDEIDQLOAT-----QTNVINNDQNAATTEEKEAAIQLOLATAV 1340
```

```
RESULT 6
Q91CH2 PRELIMINARY; PRT; 2478 AA.
AC Q91CH2:
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DE 01-OCT-2000 (TremBLrel. 15, Last annotation update)
DE FMTB.
GN FMTB.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COL;
RA Komatsuzawa H.;
RT "Staphylococcus aureus gene for affecting the methicillin
  resistance.";
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB025716; BAA93438.1;
SQ SEQUENCE 2478 AA; 262993 MW; 1C118EBE0DB03B34 CRC64;
```

Query Match 8.0%; Score 193.5; DB 2; Length 2478;
Best Local Similarity 23.6%; Pred. No. 0.02;
Matches 141; Conservative 90; Mismatches 212; Indels 155; Gaps 28;

```
Qy 4 PIGPGPIDE-----TERTPPADLSAQGLEASAAKSA-EAORIAAGAAKPKESKTDs--- 54
Db 780 PLNPDTTNEEVAERINAAKVS--GVKAIEATTTAQDLERVKKEISKIENITDSTQT 837
Qy 55 -----VERWSILRSVAVNALS-----LADKLG-----IASSNSSSTS 87
Db 838 KMDAYNEVKQAATARKQAQNASATNEEVAEADAADVAQAQKGLHDIQVVKSKQEVADT 897
Qy 88 RSADVDSSTA-TAPTPPPPTSDYKYQAQPAYDT-----IFTSTSLADIQAALVSL--- 137
```

```
Db 898 KSKVLDRKINAIOQAKVKAAD---TEVENAYNTRKQEIQNSNASTTEKQAAAYTELDTK 954
Qy 138 QDAYTNIKDTAATD-----EETAIAAEWETKNADAIVKGAQITELAKYASDNOAILDS 190
Db 955 KQEARTNL-DAANTNSDVTTAKDNTSIAAINQVQAATTKKSDAK-AEIAQKASERKTAIEA 1012
Qy 191 LGKLT-----SFDLIQTALLOSVAANNKAAELLKEMQDNPPVPGKT----- 231
Db 1013 MNDSTTEEQAAKDKVDQAVVTANADIDNAA-----ANNVDNAKTNTNEATIAATTPDA 1066
Qy 232 ---PATAQSLVDQDTATATQIEKDN-----AIGDAYFAGQNASGA 269
Db 1067 NVKPAAKQAIAQKQVQAAQETAI--DGNNGSTTEEKAQAQVQTEKTITADAAIDAHAHTNAE 1124
Qy 270 VENAKSNNSISNI-----DSAKAAIAT---AKTOIAEAOAKFFPDSPILOEA--E 313
Db 1125 VEAAK-KAAIAKIEAIOQATTTKDNKAKEIATKANERKTAIAQODITAEIAAANDVD 1183
Qy 314 QMWIOAEKDLKNIKPADG-SDVPNPGTTVGGSKQGGSSIGSIRVSMLLDDAENETASILM 372
Db 1184 NAVTOAN---SNIEAANSQNDVDQAKTTGENSIDQVTPVNNKAT-----ARNEITAILN 1235
Qy 373 SGFRQMIHMFNTENPDSQAQOELAAQARAAGAAGDSAAA-----ALADAQAQALEAA 425
Db 1236 NKLEIQATPDATDEEKQAADE--ANTENGKANQAISAATTNAQVDEKANAEEAIAINAV 1293
Qy 426 LGKAGQOQGLNALGOITASAAVVSAGVLPLOQVLRIRARYQAVVEOKLISEEDLNSAV 483
Db 1294 TPKVVVKQAQKDEIDQLOAT-----QTNVINNDQNAATTEEKEAAIQLOLATAV 1340
```

RESULT 7
O85472 PRELIMINARY; PRT; 2055 AA.

```
ID O85472:
AC O85472:
DT 01-NOV-1998 (TremBLrel. 08, Created)
DT 01-NOV-1998 (TremBLrel. 08, Last sequence update)
DT 01-MAY-2000 (TremBLrel. 13, Last annotation update)
DE EXTRACELLULAR MATRIX BINDING PROTEIN (FRAGMENT).
GN EMB.
OS Abiotrophia defectiva.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Streptococcaceae;
OC Abiotrophia.
OX NCBI_TaxID=46125;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NVS-47;
RA Manganello R., van de Rijn I.;
RT "Cloning and Characterization of emb, a Gene Encoding the Major
  Adhesin of Streptococcus defectivus.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF067776; AAD03320.1;
FT NON_TER 2055
SQ SEQUENCE 2055 AA; 215640 MW; 9699C11DDE93E2FD CRC64;
```

Query Match 7.8%; Score 188; DB 2; Length 2055;
Best Local Similarity 23.9%; Pred. No. 0.032;
Matches 124; Conservative 68; Mismatches 197; Indels 130; Gaps 24;

```
Qy 19 ADLSAQGLEASAAKSAEQAQRIAGAEA-----KPKESKTDsVERWSILRSVAVNALSAD 73
Db 1602 AKKADAIIDAATSNADVTQAQDAGKNAINAVPQTPPTAKTD-----AKNAVTAQAD 1651
Qy 74 KLGIASSNSSSTSRSDVDSTTATATPPPTSDYKYQAQPAYDTTFTSTSLADIQA- 132
Db 1652 AKKDAIENDANILTRKDAKAAKVDAAE-----TKAKNAID---AATSNADVTAK 1698
Qy 133 -----ALVSLQDAVT-----NIKDTRATDEETAI-----AAWETKNADAIVKQAQITE 176
Db 1699 QNEGTKAINDVPQTPPTAKTDKNAVDQAATDKSAIENDPALTRKDAKAAKVDNEATK 1758
```

QY 177 LAKYASDNQAIIISLGLTSFLLQTLLOSVANNKAAELLKEMQNPV--VPGKTPAI 234
DB 1759 -AKNAID-----AATSNADVTQAQKDAKNAINAVP-QIPTA 1792
QY 235 ---AQLSDQ-----TDATATQIEKD-GNAIGDAYFAGNAGAVENAKNSNI- 279
DB 1793 KTDAKNAVDAQATDKKAAIENDPALTREEKDAKAKYDA--EAKKAKDAIDAATSNADVT 1850
QY 280 SNIDSAKAAI-----ATAKTOIAEA-----QKFPDPSPILOEAFQWVIOAEKDLN 325
DB 1851 AOKDAGKDAINAVPQTPTAKTDANAVDAQATDKKSAIENDPALTREEKDAKAKYDAEA 1910
QY 326 IKPADGSDVPNPGTGVGSKOQS-SIGSI-RVSMILLDDAENETASILMSGFROMIHFN 383
DB 1911 KKAKDAIDAATSNADVTAKOTEGTOAINAVPOTPTAKTDANAVDAQATDKKNAIENDPA 1970
QY 384 TENPDQAQOELAAQARAAGDSDSAAALADAQAKALEALGKAGQOQG--TLNALGQ 441
DB 1971 LTREKDAKAKYDAEAKKAKDAID--AATSNADV-----TAKONEGTKAINDVPO 2019
QY 442 IASAAVVSAGVLPLOQVLRIRARYQAYVEQKLISEEDLN 480
DB 2020 TPTAKTDANAVD-----QAATDKKSAIENDAN 2047

RESULT 8

Q99QR6 PRELIMINARY; PRT; 2481 AA.
ID Q99QR6
AC Q99QR6
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE FMTB PROTEIN.
GN FMTB(MRP) OR SAL964.
OS Staphylococcus aureus subsp. aureus N315.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Oi Y.,
RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
RA Ogasawara N., Hayashi H., Hiramatsu K.
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus".
RL Lancet 357:1225-1240(2001).
DR EMBL; AP003136; BAB43253.1;
KW Complete proteome.
SQ SEQUENCE 2481 AA; 263767 MW; ELEAB99B81665E8 CRC64;

Query Match

Best Local Similarity 7.68; Score 182.5; DB 2; Length 2481;
Matches 136; Conservative 96; Mismatches 219; Indels 139; Gaps 26;

QY 4 PIGPGPIDE-----TERTPPADLSAOGLEASAANKSA-EAORIAGAEAKPKESKTDG--- 54
DB 763 PLNPDFTNEEVAERINAAKVS--GVKAIEATTTQDLERKNEEFKIENTIDSTQT 820
QY 55 -----VERMSILRSVAVNALMSLADKLGITASSNSSSTSRSD-----VDSTATAPT 101
DB 821 KMDAYKEVROAATARKAQNATVSNATDEEVAEANAADVAQTEGLHDIVQVKSQOEADT 880
QY 102 PP-----PPTSDDYKTOQTAYDT-----IFTSTLADTQAALVSL--- 137
DB 881 KAKVLDKINATQTAQVKPAD---TEVENAYNTRKQIIONSNASTTEKEEAAYTELDAK 937
QY 138 -ODAVTNIKDPAATD-----EETAIAAEWETKNADAIVKGAOITELAKYASDNOAILDS 190
DB 138 -ODAVTNIKDPAATD-----EETAIAAEWETKNADAIVKGAOITELAKYASDNOAILDS 190

DB 938 KOEARTNL-DAANTNSDVTTAKDNGIAAINQVQAATTKKSDAK-AEIAQKASERKTAIEA 995
QY 191 LGKLT-----SFDLLOTTALLOSVANNNKA-----AELLKEMQNPVVPKGT--PAI 234
DB 996 MNDSTTEBQQAOKDKVQAVVTANADIDNATANTDNDNAKTNTNEATTAATIPDANVKPAA 1055
QY 235 AQSIVDQTDATATQIEKDGNA-----IGDAYFAGNAGAVENAKNSN 277
DB 1056 KQATADKVAQETAIDANNSTTEEKEAKQOVOTEKTAADAAIDAASHNVEVEAAK-NA 1114
QY 278 SISI-----DSAKAIAT-----AKTOIAEAKKFPDPSPILOEA--EQWVIOAEK 321
DB 1115 EIAKIEATQIPATTTKDKNAKQAIATKANERKTAIAQTODITAEETAAANADVNDVAVTOAN- 1173
QY 322 DLKNIKPADG-SDVPNPGTGVGSKOQSSIGSIRVSMILLDDAENETASILMSGFROMIH 380
DB 1174 --SNEIANSONDVQAKTTGETSIDQVTPVNNKAT-----ARNETAILNNKLQELQA 1226
QY 381 MFNTENPDQAQOELAAQARAAGDSDSAAA-----ALADAQAKALEALGKAGQOQ 433
DB 1227 TPDATDEKQAADAE--ANTENKANQAISAATTTNAQVDEAKANAENAVTPKVVVKQ 1284
QY 434 GILNALGQIASAAVVSAGVLPLOQVLRIRARYQAYVEQKLISEEDLSAV 483
DB 1285 AAKDEIDQLQAT-----QTNVINNDQATNEEKEAAIQOLATAV 1323

RESULT 9

Q9RDQ1 PRELIMINARY; PRT; 1545 AA.
ID Q9RDQ1
AC Q9RDQ1
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE PUTATIVE SECRETED PROTEIN.
GN SC4A7.11.
OS Streptomyces coelicolor.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Streptomycetaceae; Streptomycetaceae; Streptomyces.
OX NCBI_TaxID=1902;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC Seeger K.J., Harris D.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC Cerdano A.M., Parkhill J., Barrell B.G., Rajandream M.A.;
RL Submitted (DEC-1999) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN-A3(2);
RC MEDLINE=97000351; Pubmed=8843436;
RA Redenbach M., Kieser H.M., Denapante D., Eichner A., Cullum J.,
RA Kinashi H., Hopwood D.A.;
RT "A set of ordered cosmids and a detailed genetic and physical map for
the 8 Mb Streptomyces coelicolor A3(2) chromosome";
RL Mol. Microbiol. 21:77-96(1996).
DR EMBL; AL133423; CAB62715.1;
DR InterPro; IPR000772; Ricin_B_lectin.
DR Pfam; PF00652; Ricin_B_lectin; I.
DR SMART; SM00458; RICIN; I.
SQ SEQUENCE 1545 AA; 161519 MW; 81EF325143593AEA CRC64;

Query Match

Best Local Similarity 7.58; Score 182; DB 2; Length 1545;
Matches 130; Conservative 67; Mismatches 205; Indels 126; Gaps 24;

QY 11 DETERTPPADLSA--OGLEASAANKSAEAOIAGAEAKPK---ESKTDSEVERWS----- 59
DB 741 DAKGRAETADTTATEKRGAEAKDAKAKALRDDAWDAEQKANAARAKADAKEAYAQASAS 800

QY 60 -----ILRSVNMALSLADKLGIAAS-----SNSSSTSRSDVDSTTATPTPPPT---S 107
 Db 801 EDNAQEARKAANDADAADAETAATAVAARSEADKATKAADADADAATRAEAAAKRGRADS 860
 QY 108 DDYKTOAQATVDITFTSTSLADICAAALVSLQDAVTNFKDTAAATDEETAI-----AA 158
 Db 861 DAAQAALKTADAARVATSA-----ADAIKASK-TAAATAARTAVELADDAEQHAA 910
 QY 159 EWETKNDAIKVGAQITELAKYASDNOAILDSLGKITSFDLLQTLALQSIVANNKAAELL 218
 Db 911 D-AKKEADAARAE-VTALA-----GANESTGYAVTTAAQAAVADAGNSAAQVA 955
 QY 219 KEMQD-----NPVPGKTPAIAQSLVDQTDATATQIEKDNAGNAGDFAFGONASGAVENA 273
 Db 956 APANDAIALQGLSPYV---TTDSAGLAVLTGSSKTTIAEQQQQVAEAH-----AQNAEESA 1007
 QY 274 KSNNSINIDS--AKAAIATAKTOI---AEAQKFPD-----SPILQEAQEQVIOA 319
 Db 1008 AQAQSVANAASGDSKAAATYTLAAEALGYAADARNSAKEALGYSAEAAASYATQAAQSLRTTI 1067
 QY 320 EKDLNITKPADGSDVPNPGTTVGGSKOOGSIGISIRVSMLLDDAENETASILMSGFQMI 379
 Db 1068 AYDTQATKDAAAD-----SAAQ--RAEGHAEDARDSADAAL-----1103
 QY 380 HMFNTENPDSAAQ-----QELAAQAR-AKAAGDDSA---ALADAQKALEAALCKAG 430
 Db 1104 -----DAEASARSADTAEQAEADARDADHAATGAARAEAAEAKQAYAESAAQQAEE 1155
 QY 431 QOQGLNALGQIASAAVVSAGVLPLOQVLMIRARYQAYVEQKLISEED 478
 Db 1156 QAEREANA-EQIDKGTVDQTAGIDGVYVVDHIEKIGEPVVKQSD 1202

RESULT 10
 Q9NDJ0 PRELIMINARY; PRT; 697 AA.
 AC Q9NDJ0;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE MEROZOITE SURFACE PROTEIN 3B.
 GS MSP-3B.
 OS Plasmodium vivax (strain Belem).
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporidia; Plasmodium.
 OX NCBI_TaxID=31273;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-BELEM;
 RA Galinski M.R., Al-Khedery B., Ingravallo P., Corredor-Medina C.,
 RA Barnwell J.W.;
 RT "Plasmodium vivax merozoite surface proteins-3 beta and -3 gamma share
 RT structural similarities with plasmodium vivax MSP-3 alpha and define a
 RT new gene family in Plasmodium.";
 RL submitted (OCT-1998) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AF099662; AAF78287.1;
 DR InterPro; IPR000122; Chemotaxis_transducer.
 SQ SEQUENCE 697 AA; 75685 MW; 0AF2AE9801A956A7 CRC64;

Query Match 7.5%; Score 181; DB 5; Length 697;
 Best Local Similarity 20.8%; Pred. No. 0.02;
 Matches 124; Conservative 81; Mismatches 200; Indels 192; Gaps 21;

QY 1 MWNPTGPGIDETERTPPA-----DLSAQGLEASAANKSAEAQRI 40
 Db 53 IVPNGSDVDVDEGDGEALECGNGSPKSAEPKVHAEVVKESLKSAQNAKAEAKA 112
 QY 41 AGAEAKPKESKTSVVERSI-----LRSVAVN 66
 Db 113 AKAAESAKNTLDALKVNVPTLNNVKKFAESAATEAQKQENIATEAEKKVAEANGEVV 172
 QY 67 ALMSLADKLGIASSNS-----SSSTSRSDVDSTTA-----97

Db 173 ELQKLKDEVDKAAKAKQLQKAEIAEHAQVAKTEAEKAKQDATTAKAIVAIKETGTS 232
 QY 98 -----TATPPPPSDDYKTOAQATYDIFTSTSLA-----DIOAALVSLQDAVT 142
 Db 233 KSENVTKAIDMAKKEEETKNOASTASENADRAAKAAQAEVKEIKDENKEISQLENEIT 292
 QY 143 NIKDTAATDEETAIAAEWETKNADAIVGAQI-----TELAKYASDNOAILDSLGKLTSS 196
 Db 293 KIGDIINTVKELASNAEDASKNAKKEWKAQIAAEVAKAEAKAEANANFLAEKAK--- 349
 QY 197 FDLQTLALQSIVANNKAAELLKEMODNPVPGKTPAIAQSLVDQTDATATQIEKDNAGN 256
 Db 350 ---QTA---EKIAKTSKSTEKITE-----EVRKATEFAKTAGDATQTAATE-----AA 391
 QY 257 GDYFAGQNASGAVENAKSNNSINIDSAAIATAKTO---IAEAQKFPDPSILQEE--- 311
 Db 392 GDSVSEEQKQVLESIK--OKAESALQASRKDAI--KARTEAENPLEIAKEVPRAEAAKEEA 449
 QY 312 -----AEQMVIOAEKDLKNIKPADGSDVPNPGTTVGGSKOOGSIGISIRVSMLLDDAEN 365
 Db 450 QKAATSADAETKALTAIEVKNKSDAS-----EN 478
 QY 366 ETASILMSGFQMIHFWNTE-NPDSQAQOELAAQARAARAKAAGDSDSAALADAQKALEA 424
 Db 479 EKKKI-----ETEANATAGEAQAARAKAFAKAAADAKDTNEAVTLAVAKEKVEK 526
 QY 425 ALGKAGQOQGLNALGQIASAAVVS---GVLPLQOVLMIWIRARYQAYVEQKLISEE 477
 Db 527 AL-KAAKEAKKAN---EKASYALIRTKQYALEPLE---ITSEAGYNITEKEEQVKEE 577
 RESULT 11
 Q99QY4 PRELIMINARY; PRT; 2271 AA.
 AC Q99QY4;
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE SA2447 PROTEIN.
 GN SA2447.
 OS Staphylococcus aureus subsp. aureus N315.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Staphylococcus.
 OX NCBI_TaxID=158879;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
 RA Cui L., Oguchi A., Aoki K.I., Nagai Y., Lian J., Ito T., Kanamori M.,
 RA Matsumaru H., Maruyama A., Murakami H., Hosoyama A., Mizutani-Ui Y.,
 RA Takahashi N.K., Sawano T., Inoue R.I., Kaito C., Sekimizu K.,
 RA Hirakawa H., Kuhara S., Goto S., Yabuzaki J., Kanehisa M.,
 RA Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T., Hattori M.,
 RA Ogasawara N., Hayashi H., Hiramatsu K.;
 RT "Whole genome sequencing of methicillin-resistant Staphylococcus
 RT aureus.";
 RT Lancet 357:1225-1240(2001).
 RL EMBL; AP003138; BAB43752.1;
 DR Complete proteome.
 KW SEQUENCE 2271 AA; 227847 MW; 7C2A7040D6C8289D CRC64;

Query Match 7.5%; Score 180; DB 2; Length 2271;
 Best Local Similarity 18.9%; Pred. No. 0.1;
 Matches 79; Conservative 96; Mismatches 187; Indels 56; Gaps 12;

QY 22 SAQGLEASAANKSAEAQRIAGAEAKPKESKTSVERWSILRSVNMALSLADKLGIASSN 81
 Db 1516 SASASESSSTSDSTSDSTASMQSESQSTASLSLSTSTSNRMT-----IASLS 1570
 QY 82 SSSSTSRSDVDSTTATPTPPPTSDDYKTOAQAT-AYDTIFTSTSLADIQAALVSLQDA 140
 Db 1571 TSVSTSEGSTSESTSESDSTSLSDSQSTSRSTASGSASTSTSTSDSRSTASTS-- 1628

[illegible]

```

Query Match      7.2% Score 174.5; DB 2; Length 1795;
Best Local Similarity 22.4%; Pred. No. 0.15;
Matches 132; Conservative 93; Mismatches 228; Indels 135; Gaps 26;
```

QY	4	PIGPGPIDE-----TERTPPADLSAOGLEASAANKS-AEAGIAGAAGAKPKESKTD:-	--- 54
DG	783	PLNPDTTNEVEVAEIERINAAKV:-GVKAIEATTTAQDLVRVKNKEISKIENTIDSTQT	840
QY	55	-----VERWSILRSVNAALMS-----LADKLG-----IASSNSSSTS	87
DG	841	KMDAYNEVKQAATAARKAQNATVSNATNEEVAEADAEEAAQKGLDHDIQVVKSQEVA	900
QY	88	RSADVDSSTA-TATPPPPPSDDKYTKTAQATYDT-----IFTSTSLADIQAALVSL---	137
DG	901	KSKVLDKINAIQTKAVKVPRAAD---TEVENATYRKRQEIQNSNASTTEEKQAAYTELDTK	957
QY	138	-QDAVTNI-----KDTAATDEETAIAAEWETKNADATKVGAQITELAKYASDNQAILDSL	191
DG	958	KQEARTNLDAANTNSAVTTAKDNGIAAINVOQATTKKSDAK-AELAQKASERKTATEAM	1016
QY	192	GKLT-----SFDLLQTLALQSVAANNKA-----AELLKEMODPNVPVGKT--PAIA	235
DG	1017	NDSYTEEQQAAKEVDQAVLSANADIDNAEAANTDVDNAKTTNEATTAITPDANKVPOAK	1076
QY	236	QSLVDQDTDATATQLEKDGNA-----IGDAYFAGQNASGAVENAKSNS	278
DG	1077	QAIAADKVOAQEKAIDANNGSTTEEKAACKQOVQTEKTTADADAIDAATHNAEVEAAK-NAE	1135
QY	279	ISNT-----DSAKAAAIAT-----AKTOIAEAKKFDPDSPILQEA---BQMVIQAEKD	322
DG	1136	IAKTEAIOAPPTTKDNKAELATKANERKTAIAOTODITABEIAAANADVNAVTOYTON--	1193

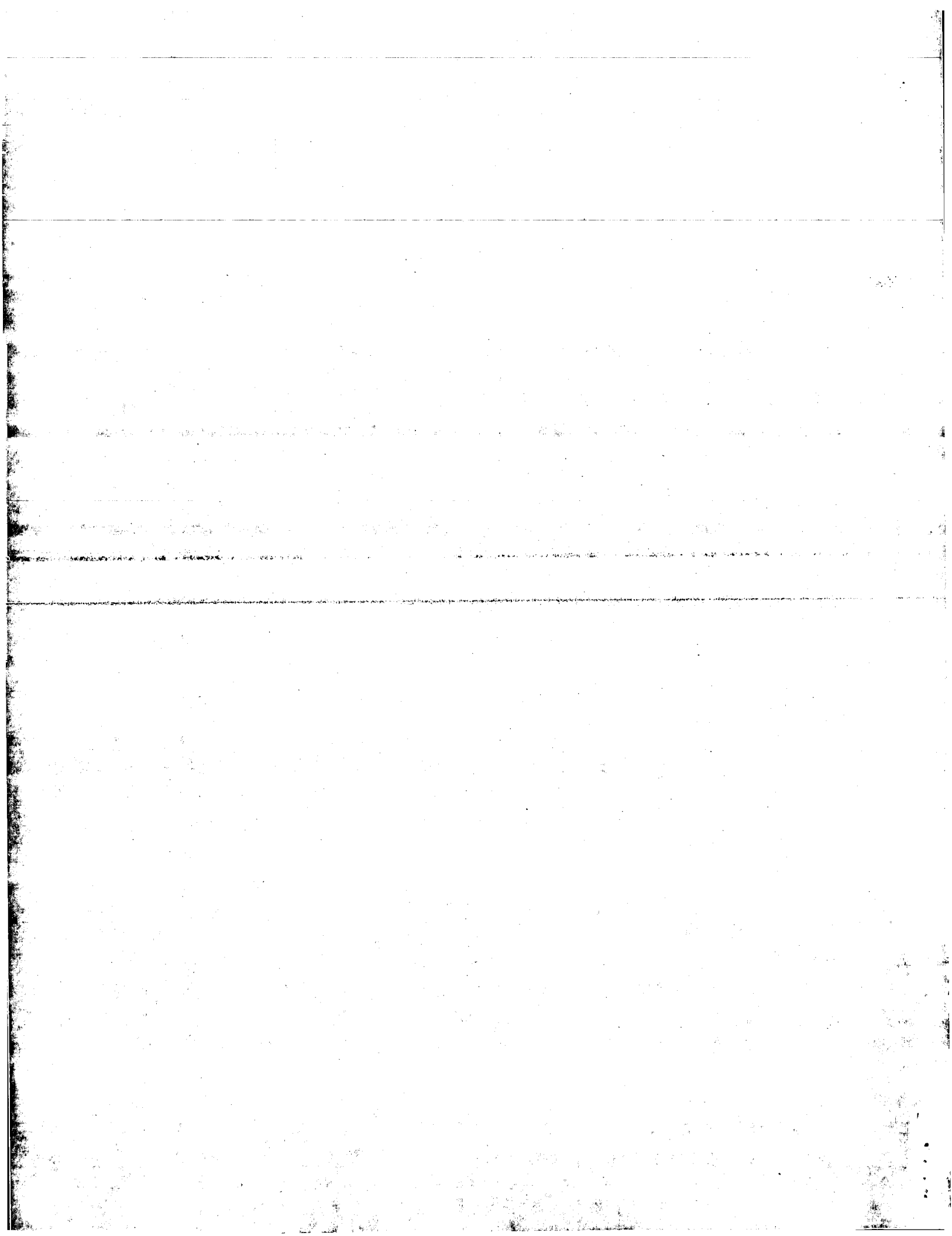
Query Match	7.2%;	Score 174.5;	DB 2;	Length 1795;
Best Local Similarity	22.4%;	Pred. No. 0.15;		
Matches 132;	Conservative 93;	Mismatches 228;	Indels 135;	Gaps
QY	4	PICPGPTIDE-----TERTPPADLSAAGLEASAAKSA-EAORIAGAEAKPKESKTDG---	54	
Db	783	PLNPDTTNEVEAEIERINAAKVS--GVKAIEATTAAOGLERVNEELSKIENTIDSTQTT	840	
QY	55	-----VERWSILRSAVNALMS-----LADKLG-----IASSSSSSSTS	87	
Db	841	KMDAYNEVQOAAATARKAAQNAATVSNATNEEVEAEADAEEAAQKGLHDIQVVKSKOEVA	900	
QY	88	RSADVDTA-TAPTPPPPTSDDYTKTAQATYDT-----IFTSTLSADIQAALVSL---	137	
Db	901	KSVLDKINAIQTQAKVKPAAD---TEVENAYNTRKBQIENSNASTEETKQAAAYTE	957	
QY	138	-QDAVTNI-----KDTAAATDEETAATAAEWEKTKNADAIKVGQAQITELAKARYASD	191	
Db	958	KQEARTNLDAANTNSAVTTAKDNGIAAINOVQAAATTKKSDAK-AEIAQKASERKTA	1016	
QY	192	GKIT-----SFDLQITALLQSVANNKA-----BELLKEMQDNPVPVPGKT--PAIA	235	
Db	1017	NDSTTEEQAAAKEKVDQAVLSANADIDNAANTVDNNAKTTNEATIAITPDANVKPQAK	1076	
QY	236	QSLVDDOTDATABQIEKDNGNA-----IGDAYFAGQNASGAVENAKSNNS	278	
Db	1077	QATADKQVQEKAIIDANNNGSTTEEKAANKQOVQTEKTTADDAIDAHTNAEVEAAK-NAE	1135	
QY	279	ISNI-----DSAKAAIAT-----AKQTQIAEAKKFPDPSPILOEA--EQMWIQAEKD	322	
Db	1136	IAKIEAIQTATTTKNAKEAIAIKANERKRTAIAOTDITAEIEAANAADVNAVYTOAN---	1193	

QY 323 LKNKPADG--SDVNPNGTGVGSKQGGSSIGSIRVSMILLDDAENETASILMSGFROMIHM 381
 Db 1194 -SNTAANSQNDVQAKTGTGTSIDQVTPVKNKAT-----ARNETAILNKNKLOEIQT 1247
 QY 382 FNTENPDQAAQOEL-----AAQARAARAAAGDSAAAALADAQAALKAALGKAGQOQGI 435
 Db 1248 PDATDEKQADAEANTENGKALQATAA-ATTNAQVDEAKTNAEAAINAVTPKVVKKQTA 1306
 QY 436 LNALGQTASAAVVSAGVLPQQVLWIRARYQAYVEQKLISEEDLNSAV 483
 Db 1307 KDEIDQLOAT-----QTNWINNDQNATTEKEAAIQOLATAV 1343
 RESULT 14
 Q9BIT0 PRELIMINARY; PRT; 2016 AA.
 AC Q9BIT0;
 DT 01-JUN-2001 (TREMELrel. 17, Created)
 DT 01-JUN-2001 (TREMELrel. 17, Last sequence update)
 DE FIBROIN 3 (FRAGMENT).
 OS Plectreureys tristis (Spider).
 OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Araneae;
 OC Araneomorphae; Haplogynae; Plectreureidae; Plectreureys.
 OX NCBI_TaxID=33319;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21179804; PubMed=11283372;
 RA Gatesy J., Hayashi C., Motriuk D., Woods J., Lewis R.;
 RT "Extreme Diversity, Conservation, and Convergence of Spider Silk
 Fibrin Sequences";
 RL Science. 291:2603-2605(2001).
 DR EMBL; AF350283; AAK30612.1;
 FT NON_TER
 SQ SEQUENCE 2016 AA; 196162 MW; F53A47DB185826F1 CRC64;
 Query Match 7.2%; Score 173; DB 5; Length 2016;
 Best Local Similarity 21.5%; Pred. No. 0.21;
 Matches 103; Conservative 83; Mismatches 212; Indels 88; Gaps 16;
 QY 19 ADLSAQGLEASAANKSAEQAQRIAGAEAKPKESKTDVSVERMSILRSVAVNMLSLADKLGIA 78
 Db 1241 ASASAYELEFSDSSSSAVAVASSQGSYDTSDD-----FSSSSAAAAAASAYESFLD 1296
 QY 79 SSSSSSTSRADVDSTATATPTPPPTSDDYKTAQATAYDTIPTSTSLADIAALVSLQ 138
 Db 1297 ASSSSSSAAAASSQGSYETSSDLVSSAAAAAASASAYSQFLDASSSSNAATSSR 1356
 QY 139 ---DAVTNIKDTATDETAIAAEWETKNADAIVKGAQITELAKYASDNQAILDLSGLK 194
 Db 1357 QSSYDTSDFSSASIAAAAAAASAYESQFSDA-----SSSSNAAAAASSQGSYD----- 1407
 QY 195 TSFLLQTA-----LLQSVANNKRAELKEMQNPVPGKTPAIAQSLVD-QTDAT 245
 Db 1408 TSSDLVSAASASAYESQFLDASSSSNA-----TTSSQGSYDTSDFSS 1452
 QY 246 ATQIEKQCNAGDAY-----FAGQNASGAVENAKSNNSISN--IDSAAKAIATAKTQI 296
 Db 1453 SASIAAAAAAASAYESQFSDASSSSNAAAAAAASQGSYDTSDDLVSASSAAAAAASASS 1512
 QY 297 AEAKKFPDPTLOPA-----EQMVIOAEKDLKNIPKADG-----SDVNPNG 338
 Db 1513 YESQ--FSDASSSSNAAAAAAASQGSYDTSDDLVSASSAAAAAASAYESQFSDASSSR 1570
 QY 339 TTGVGSKQGGSSIGSIRVSMILLDDAENETASILMSGFR-QMHHFNTENPDQAAQOE-- 395
 Db 1571 NAAAAAASQGSYDPT--SSDLVSSAAAAAASAYESQFLDASSSSNAATTSQGS 1628
 QY 396 -----LAAQARAARAAAGDSAAAALADAQAALKAALGKAGQOQ-----ILNALGQI 442
 Db 1629 YDTSDFSSASIAAAVAASAYESQFSDASSSSKAAAAAASQGSYDTSDDLVSASSAA 1688

QY 443 ASAAVVSA 450
 Db 1689 AAAASASS 1696
 RESULT 15
 Q9HS86 PRELIMINARY; PRT; 627 AA.
 ID Q9HS86
 AC Q9HS86;
 DT 01-MAR-2001 (TREMELrel. 16, Created)
 DT 01-MAR-2001 (TREMELrel. 16, Last sequence update)
 DE HTR14 TRANSDUCER.
 GN HTR14 OR VNG0355G.
 OS Halobacterium sp. (strain NRC-1).
 OC Archaea; Euryarchaeota; Halobacteriales; Halobacteriaceae;
 OC Halobacterium.
 OX NCBI_TaxID=64091;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20504483; PubMed=11016950;
 RA Ng W.V., Kennedy S.P., Mahairas G.G., Berquist B., Pan M.,
 RA Shukla H.D., Lasky S.R., Baliga N.S., Thorsson V., Sbrogha J.,
 RA Swartzell S., Weir D., Hall J., Dahl T.A., Welti R., Goo Y.A.,
 RA Leithausser B., Keller K., Cruz R., Danson M.J., Hough D.W.,
 RA Maddocks D.G., Jablonski P.E., Krebs M.P., Angevine C.M., Dale H.,
 RA Isenbarger T.A., Peck R.F., Pohlschroder M., Spudich J.L., Jung K.-H.,
 RA Alam M., Freitas T., Hou S., Daniels C.J., Dennis P.P., Omer A.D.,
 RA Ebhardt H., Lowe T.M., Liang P., Riley M., Hood L., Dassarma S.;
 RT "Genome sequence of Halobacterium species NRC-1";
 RL Proc. Natl. Acad. Sci. U.S.A. 97:12176-12181(2000).
 DR EMBL; AE004994; AAG18922.1;
 DR InterPro; IPR000122; Chemotaxis_transducer.
 DR InterPro; IPR000658; DUF5.
 DR InterPro; IPR003660; HAMP.
 DR Pfam; PF00672; DUF5; 1.
 DR Pfam; PF00015; MCPsignal; 1.
 DR PRINTS; PR00260; CHEMTRNSDUCR.
 DR SMART; SM00304; HAMP; 2.
 DR SMART; SM00283; MA; 1.
 KW Complete proteome.
 SQ SEQUENCE 627 AA; 65655 MW; E5F2867FA6CA2A75 CRC64;
 Query Match 7.1%; Score 171; DB 1; Length 627;
 Best Local Similarity 23.4%; Pred. No. 0.063;
 Matches 120; Conservative 87; Mismatches 184; Indels 122; Gaps 27;
 QY 17 PPADLSAQGLEASAANKSAEQAQRIAGAEAKPK-----ESKTDVSVERMS-ILRSVAVNMLSL 71
 Db 145 PALDESVPGAFGESITEMADSLAYTALEDKTAELHQOAELEROSEQURLVLDALSEA 204
 QY 72 ADKLGIASSSSSTSRSDVDSTATATPTPPPTSDDYKTAQ-----QATYDTIFTS 124
 Db 205 TD-----AARAGDL---TATVDAALDVTDDHRAAVEDFNQLETLADTISDI 249
 QY 125 TSLADIAQAL-----VSLQDAVTNIKDTA--ATDETAIAAEWETKNADA 167
 Db 250 QSFSDAVLAVSRITDERVDVADRSAAVSESVEIADGANQOTNQLNIAAEMDVTSATV 309
 QY 168 IKVGAQITELAKYASDNQAIL-----DSLQKL-TSFDLQTLQSVANNKAAELKEMQ 222
 Db 310 EETAAASANDAKTA-----QAAADRGEDGGEVEETIEALRALREQSQAETVESLAEEVE 366
 QY 223 DNPVPGKTPAIAQSLVDOTDATA--TQIE-----KQNGAIGDAYFAGQNASGAVENAKS 275
 Db 367 R---IDGIT-ALIEDIAEETNMLNALNASIEAARTGSDG--GFAVVADEVKDLAETREQ 420
 QY 276 NNSTSNTDSA-----KAAATATAKTQIAEAKKFPDPS-ILQAEQMVIOAEKDLKNKP 328
 Db 421 AADISEIVDAVTEKAEDASIAIGEVD-AEVEKTKITAEGLVRDFEIV----DEVANVNH 475
 QY 329 A--DGSVDPNPG-----TTVGSKQGGSSI-----GSIRVSMILLDD 362

Db	476	AVQEISDATDQGAOSVTDVYGMVEVASVEETAESDTVADNAAEQTDATDEVADQMD	535
QY	363	AENETAST--LMSGFROMIHWFNTENPDSQAAQOELAAQAPAAK--AAGDDSAALADA	418
Db	536	LAEQTAALAGMLDDF-----TVPADAGTADQSVADDSPTAQPPAADDEPAAAVVDQ	587
QY	419	QKALEAALGRAGQOQOQILNALGQIASAAVYSAG	451
Db	588	QPASDAE-----DEEGVPDSGGE--SVAVSDGG	613

Search completed: February 7, 2002, 21:41:15
Job time: 4644 sec



GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:34:36 ; Search time 137.02 Seconds

(without alignments)
250.299 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261

Sequence: 1 MNPVGPIDETERTPPAD.....SAAVVSAGVLPQQVLWIRA 463

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 522463 seqs, 74073290 residues

Total number of hits satisfying chosen parameters: 522463

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_1101.*
1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2241	99.1	490	21	AA1987
2	2188	96.8	651	21	AA1984
3	2184	96.6	452	21	AA1985
4	2162	95.6	478	20	AA1983
5	1850	81.8	583	21	AA1985
6	281.5	12.5	350	20	AA1987
7	231.5	10.2	361	21	AA1985
8	231.5	10.2	361	22	AA1985
9	181.5	8.0	1822	13	AA1985
10	164.5	7.3	331	20	AA1985
11	153.5	6.8	1237	21	AA1985

12	153	6.8	600	22	AA1984
13	152	6.7	1561	17	AA1985
14	150.5	6.7	643	16	AA1984
15	148	6.5	729	22	AA1984
16	148	6.5	881	22	AA1984
17	147.5	6.5	1095	22	AA1984
18	146.5	6.5	571	21	AA1984
19	143	6.3	1566	16	AA1984
20	142	6.3	1607	19	AA1984
21	140.5	6.2	564	16	AA1984
22	140	6.2	876	22	AA1984
23	139	6.1	5024	22	AA1984
24	138	6.1	1576	21	AA1984
25	138	6.1	1576	21	AA1984
26	138	6.1	1584	21	AA1984
27	138	6.1	1609	19	AA1984
28	138	6.1	1609	21	AA1984
29	138	6.1	1609	21	AA1984
30	138	6.1	1617	21	AA1984
31	138	6.1	2742	21	AA1984
32	138	6.1	2842	15	AA1984
33	138	6.1	2843	13	AA1984
34	138	6.1	2843	13	AA1984
35	138	6.1	2843	16	AA1984
36	138	6.1	2843	18	AA1984
37	138	6.1	2843	19	AA1984
38	138	6.1	2843	19	AA1984
39	138	6.1	2843	19	AA1984
40	138	6.1	2843	21	AA1984
41	138	6.1	2860	15	AA1984
42	138	6.1	2973	19	AA1984
43	138	6.1	2973	21	AA1984
44	138	6.1	2973	22	AA1984
45	137.5	6.1	2453	21	AA1984

ALIGNMENTS

RESULT 1

AA1987
ID AA1987 standard; Protein; 490 AA.

AC AA1987

DT 26-MAR-2001 (first entry)

DE C. pneumoniae 76 kba protein truncation mutant fusion protein.

XX 76 kba protein; bactericidal; diagnosis; prevention;

KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;

KW acute respiratory disease; cough; sore throat; hoarseness; fever;

KW vaccine; immunisation; treatment; fusion protein; truncation mutant;

KW munein.

XX Chlamydia pneumoniae.

OS Synthetic.

EH Key Location/Qualifiers

FT Region 453..490

FT /note= "This part of the sequence is unrelated to the C. pneumoniae 76 kba protein"

FT WO2000066739-A2.

PD 09-NOV-2000.

XX 03-MAY-2000; 2000WO-CA00511.

XX 03-MAY-1999; 99US-0132270.

XX 30-JUN-1999; 99US-0141276.

XX (AVET) AVENTIS PASTEUR LTD.

C glutamicum prote
S. mutans antigen
Trypanosoma cruzi
Mycobacterium tube
S. cerevisiae apopt
S. epidermidis ope
Arabidopsis thalia
Immunodominant ant
Mouse laminin Gl c
Trypanosoma cruzi
C glutamicum prote
S. epidermidis ope
Human laminin 2 ma
Human laminin 8 po
Human laminin 2 ga
Human laminin 2 ga
Human APC protein
Adenomatous polypo
APC gene product 1
Adenomatous polypo
Adenomatous polypo
Human adenomatous
Human APC protein
Human APC protein
Human adenomatous
Adenomatous polypo
Human APC protein.
Protein used in ca
Transcriptional ac
HNRCP protein sequ

XX Murdin AD, Oomen RP, Wang J, Dunn P;
 XX WPI: 2000-687542/67.
 DR N-PSDB; AAD02066.
 XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 PT useful for vaccinating against Chlamydia infections -
 XX Claim 33; Fig 3; 112pp; English.
 XX The present sequence is a fusion protein comprising a
 CC truncated Chlamydia pneumoniae 76 kDa protein and vector-encoded
 CC residues. C. pneumoniae 76 kDa protein is used in the
 CC diagnosis, prevention and treatment of C. pneumoniae
 CC infections (e.g. pneumonia, upper respiratory
 CC tract disease, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX Sequence 490 AA;
 SQ
 Query Match 99.1%; Score 2241; DB 21; Length 490;
 Best Local Similarity 99.1%; Pred. No. 5.5e-143;
 Matches 459; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
 QY 1 MNPPIGPGPIDETPTPADLSAAGLEASANKSAEQAQRIAGAEAKPKESKTDVSVERWSI 60
 Db 1 mvpnpigppidetertppadlsaggleasaanksaesaaqriagaekpketsdsvrws 60
 QY 61 LRSVAVNLSLADKLGIASSNSSSTSRSDVDSTTATAPPPPTSDDYKTQATAYDT 120
 Db 61 lrsavnlmsladdklgiassnssstsrsvdvtattappppttdyktqataydt 120
 QY 121 IFTSTSLADIQAALVSLQDAVNTIKDTAATDEETAIKAEWETKNADAKVGAQITELAKY 180
 Db 121 iftstsladiqaalvslqdvntikdtaataetiaaewetknadavkvgagitelaky 180
 QY 181 ASDNQAILDSLGLTSPDLLQATALLQSVANNKAAELLKEMQDNPVPGKTPALQAQSLVD 240
 Db 181 asdnqaildslgltspdllqatallqsvannnkaaellkemqdpvpgktpalqaqslvd 240
 QY 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKNSNISNIDSAKAAIATAKTQIAEAQ 300
 Db 241 qtdatatqiekdgnaigdayfagqnasgavenaknsnnsidnsakaaiaataktqiaeq 300
 QY 361 DDAENETASILMSGFROMIHMFNTENPDSQAQQLAAQARAAGDSDSAAALADAQK 420
 Db 361 ddaenetasilmsgfrimhfntenpdsqaqqelaaraagdsdaaaladaqk 420
 QY 421 ALEAALGKAGQOQGLINAGIASNAVVSAGVLPLOQVLIWIRA 463
 Db 421 aleaalgkagqqqglinagiasaavvsagvlploqvliwira 463
 RESULT 2
 ID AAY71954 standard; Protein: 651 AA.
 XX AAY71954;
 AC AAY71954;
 XX 26-MAR-2001 (first entry)
 DT Chlamydia pneumoniae 76 kDa full-length protein.
 DE Chlamydia pneumoniae 76 kDa full-length protein.
 XX 76 kDa protein; bactericidal; diagnosis; prevention; treatment;
 KW

KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;
 KW vaccine; immunisation.
 XX Chlamydia pneumoniae.
 OS WO200066739-A2.
 PN 09-NOV-2000.
 PD 03-MAY-2000; 2000WO-CA00511.
 PF 03-MAY-1999; 99US-0132270.
 PR 30-JUN-1999; 99US-0141276.
 XX (AVET) AVENTIS PASTEUR LTD.
 PA Murdin AD, Oomen RP, Wang J, Dunn P;
 XX WPI: 2000-687542/67.
 DR N-PSDB; AAD02063.
 XX Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 PT useful for vaccinating against Chlamydia infections -
 XX Claim 16a; Fig 1; 112pp; English.
 XX The present sequence is Chlamydia pneumoniae full-length
 CC 76 kDa protein. C. pneumoniae 76 kDa protein
 CC is used in the diagnosis, prevention and treatment
 CC of C. pneumoniae infections (e.g. pneumonia, upper respiratory
 CC tract disease, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX Sequence 651 AA;
 SQ
 Query Match 96.8%; Score 2188; DB 21; Length 651;
 Best Local Similarity 98.9%; Pred. No. 2.9e-139;
 Matches 449; Conservative 1; Mismatches 4; Indels 0; Gaps 0;
 QY 1 MNPPIGPGPIDETPTPADLSAAGLEASANKSAEQAQRIAGAEAKPKESKTDVSVERWSI 60
 Db 1 mvpnpigppidetertppadlsaggleasaanksaesaaqriagaekpketsdsvrws 60
 QY 61 LRSVAVNLSLADKLGIASSNSSSTSRSDVDSTTATAPPPPTSDDYKTQATAYDT 120
 Db 61 lrsavnlmsladdklgiassnssstsrsvdvtattappppttdyktqataydt 120
 QY 121 IFTSTSLADIQAALVSLQDAVNTIKDTAATDEETAIKAEWETKNADAKVGAQITELAKY 180
 Db 121 iftstsladiqaalvslqdvntikdtaataetiaaewetknadavkvgagitelaky 180
 QY 181 ASDNQAILDSLGLTSPDLLQATALLQSVANNKAAELLKEMQDNPVPGKTPALQAQSLVD 240
 Db 181 asdnqaildslgltspdllqatallqsvannnkaaellkemqdpvpgktpalqaqslvd 240
 QY 241 QTDATATQIEKDGNAIGDAYFAGQNASGAVENAKNSNISNIDSAKAAIATAKTQIAEAQ 300
 Db 241 qtdatatqiekdgnaigdayfagqnasgavenaknsnnsidnsakaaiaataktqiaeq 300
 QY 301 KFPDPSPIQAEQMVTOAEKDLKNIKPADGSDVPNPPTTVGGSKQGSSIGSTRVSMML 360
 Db 301 kfpdspilqaeqmvtoaeqdlknikpadgsdvpnppttvggskqgssigstrvsmll 360
 QY 361 DDAENETASILMSGFROMIHMFNTENPDSQAQQLAAQARAAGDSDSAAALADAQK 420
 Db 361 ddaenetasilmsgfrimhfntenpdsqaqqelaaraagdsdaaaladaqk 420
 QY 421 ALEAALGKAGQOQGLINAGIASNAVVSAGVLP 454

```

|||||
Db 421 aleaalgkagggqglnalgiaseavvsagvpp 454

RESULT 3
AAY71956
ID AAY71956 standard; Protein: 452 AA.
XX
AC AAY71956;
XX
DT 26-MAR-2001 (first entry)
XX
DE 3'-truncated Chlamydia pneumoniae 76 kDa protein.
XX
KW 76 kDa protein; bactericidal; diagnosis; prevention;
KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
KW acute respiratory disease; cough; sore throat; hoarseness; fever;
KW vaccine; immunisation; treatment; truncation mutant; mutein.
XX
OS Chlamydia pneumoniae.
OS Synthetic.
PN WO200066739-A2.
XX
PD 09-NOV-2000.
XX
PF 03-MAY-2000; 2000WO-CA00511.
XX
PR 03-MAY-1999; 99US-0132270.
PR 30-JUN-1999; 99US-0141276.
XX
PA (AVET ) AVENTIS PASTEUR LTD.
XX
PI Murdin AD, Oomen RP, Wang J, Dunn P;
XX
WPI: 2000-687542/67.
DR N-PSDB: AAD02065.
XX
Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
PT useful for vaccinating against Chlamydia infections -
XX
Claim 16c; Page 104-106; 112pp; English.
XX
The present sequence is 3'-truncated
CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
CC protein is used in the
CC diagnosis, prevention and treatment of C. pneumoniae
CC infections (e.g. pneumonia, sinusitis and acute respiratory
CC tract disease, bronchitis, sinusitis and acute respiratory
CC disease such as cough, sore throat, hoarseness, fever; and
CC abnormal chest sounds on auscultation). C. pneumoniae sequence
CC is also used as vaccines for immunising humans against diseases
CC caused by C. pneumoniae.
XX
SQ Sequence 452 AA;

Query Match 96.6%; Score 2184; DB 21; Length 452;
Best Local Similarity 99.1%; Pred. No. 3.4e-139;
Matches 448; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MWNPTGPGPIDETPTPPADLSAQGLEASANKSAEORAGAEAKPKESKTSVERWSI 60
Db 1 mwnpgppideterppadlsaggleasaanksaeariagaeakpkeskdsvrws1 60
QY 61 LRSVNALMSLADKLGIASSSSSTSRSDVDSSTATPTPPPTSDYKTAQATYDT 120
Db 61 lrsavnlmsladvkligiassssstsrsvdvtatptppptfddyktaqatyd 120
QY 121 IFTSTSLADIQAALVSLQDAVTNKTATDEETAIAAEWETKNADAIVGAQITELAKY 180
Db 121 iftstsladiqaalvslqdvatnktatdeetaiaaewetknadvkvgaqitelaky 180

181 ASDNOAILDLSGLKTSFDLLOTFALLOSVANNKKAELLKEMQDNPVPGKTPATAQSIVD 240
|||
Db 181 asdnqailds1gktsfdllqallqsvannnkkaellkemdnpvpgktpataqsivd 240

241 QTDATATQIEKDGNAGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTQIAEAQ 300
|||
Db 241 qtdatatqiekdnairdayfagqnasgavenaksnnsidsakaiaiataktqiaeaq 300

301 KKFPDSPILQAEQWVIOAEKDLNKNRPADGSDVPNPGTTVGGSKQOGSSIGSRVSMLL 360
|||
Db 301 kkfpdspilqaeqwmviquekdlknkpadgsdvpnpgttvggskqgssigsrvsmll 360

361 DDAENETASILMSGFROMIHMFNTENPDSSQAQOELAAOARAAKAGDSDSAAALADAOK 420
|||
Db 361 ddaenetasilmsgfrqmihmfntenpdsqaqqelaaraakaagddsaaladaok 420

421 ALEAALGKAGQOQGLNALGOITASAAVVSAGV 452
|||
Db 421 aleaalgkagqgglnalgiaseavvsagv 452

RESULT 4
AAY35358
ID AAY35358 standard; Protein: 478 AA.
XX
AC AAY35358;
XX
DT 13-SEP-1999 (first entry)
XX
DE Chlamydia pneumoniae surface exposed polypeptide.
XX
KW Respiratory disease; pneumonia; bronchitis; heart disease; sarcoidosis;
KW sinusitis; purulent otitis media; erythema nodosum; pharyngitis;
KW vaccine; neutralising epitope.
XX
OS Chlamydia pneumoniae.
XX
WQ9927105-A2.
XX
03-JUN-1999.
XX
20-NOV-1998; 98WO-IB01890.
XX
04-NOV-1998; 98US-0107078.
PR 21-NOV-1997; 97FR-0014673.
XX
(GEST ) GENSET.
XX
Griffais R;
XX
WPI: 1999-357842/30.
XX
Genome sequence of Chlamydia pneumoniae
XX
Page 1156; Disclosure; 1912pp; English.
XX
AAY34584-Y35879 represent the proteins encoded by all the open reading
frames in the complete genome (see AAX91990) of Chlamydia pneumoniae.
CC C. pneumoniae causes respiratory disease such as pneumonia and
CC bronchitis and is thought to be a contributing factor in heart
CC disease, sarcoidosis, sinusitis, purulent otitis media, erythema
CC nodosum or pharyngitis. The polypeptides encoded by the open reading
CC frames of the C. pneumoniae genome (see AAY34584-Y35879) can be used in
CC immunogenic compositions as vaccines. Vectors containing C. pneumoniae
CC nucleotide sequences can also be used as immunogenic compositions,
CC especially where the vector directs the expression of a neutralising
CC epitope of C. pneumoniae.
XX
SQ Sequence 478 AA;

Query Match 95.6%; Score 2162; DB 20; Length 478;
Best Local Similarity 99.1%; Pred. No. 1.1e-137;

```

Matches 443; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 MWNPIGPGPIDETERTPPADLSAAGLEASANKSAEQRAGAEAKPKESKIDSVVERWSI 60
 Db 3 mwnpiggpiddetertppadlsaggleasaanksaegrIagaeakpkesktidsverwsi 62
 QY 61 LRSVAVNLMADKLGIASSNSSSTSRSDVDSTTATAPPPPTSDVKYKTAQAYDT 120
 Db 63 lrsavnaImadklgiassnssstsrSadvdsttatapppttdfdyktqagaydt 122
 QY 121 IFTSLADIQAALVSLQDVTNIDTAATDEETAIAAEWETKNADAIVKGAQITELAKY 180
 Db 123 iftstsladiqaalvslqdvtnikdtaateetaiaaewetknadvkvgaqitelaky 182
 QY 181 ASDNQAIILDSLGKLTSLFLLQVANNKAAELLKEMQDNPVPGKTPAIAQSLVD 240
 Db 183 asdnqaiildslgkltslfllqavannnkaaellkemqdnvpvpgktpaiaqslvd 242
 QY 241 OTDATATQIEKDGNAIGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAQ 300
 Db 243 qtdatatqiekdgndairdayfagqnasgavenaksnnsisnidsakaaiaatakqiaeaq 302
 QY 301 KKFDPSPILQAEQMVIOAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLL 360
 Db 303 kkfpdspilqaeqmviqaeKdlknikpadgsdvpnpgtttvggskqgssigsirvsmll 362
 QY 361 DDAENETASILMSGFROMIHMFNTENPDSSQAQELAAQARAANKAGDDSAALADAQK 420
 Db 363 ddaenetasilmsgfrmiHmfntenpdsgaaqelaaqaraaKaagddsaaladaqk 422
 QY 421 ALEAALGKAGQOQILNALGQTASAAV 447
 Db 423 aleaalgkagqggilnalqiasaav 449

RESULT 5

AAV71955
 ID AAV71955 standard; Protein; 583 AA.
 AC AAV71955;
 XX
 XX
 DT 26-MAR-2001 (first entry)
 DE
 XX
 XX
 KW 76 kDa protein; bactericidal; diagnosis; prevention;
 KW pneumonia; upper respiratory tract disease; bronchitis; sinusitis;
 KW acute respiratory disease; cough; sore throat; hoarseness; fever;
 KW vaccine; immunisation; treatment; truncation mutant; mutein.
 XX
 OS Chlamydia pneumoniae.
 OS Synthetic.

FH Key Location/Qualifiers
 FT Misc-difference 497 /note= "Encoded by ACT"

XX
 PN WO2000066739-A2.
 XX
 PD 09-NOV-2000.
 XX
 PF 03-MAY-2000; 2000WO-CA00511.
 XX
 XX 03-MAY-1999; 99US-0132270.
 PR 30-JUN-1999; 99US-0141276.
 XX
 XX (AVET) AVENTIS PASTEUR LTD.
 PA
 XX Murdin AD, Oomen RP, Wang J, Dunn P;
 PI
 XX WPI: 2000-687542/67.
 DR N-PSDB; AAD02064.
 XX

PT Nucleic acids encoding a 76 kDa protein from Chlamydia pneumoniae,
 PT useful for vaccinating against Chlamydia infections -
 XX
 PS Claim 16b; Page 100-102; 112pp; English.
 XX
 CC The present sequence is 5'-truncated
 CC Chlamydia pneumoniae 76 kDa protein. C. pneumoniae 76 kDa
 CC protein is used in the
 CC diagnosis, prevention and treatment of C. pneumoniae
 CC infections (e.g. pneumonia, upper respiratory
 CC tract disease, bronchitis, sinusitis and acute respiratory
 CC disease such as cough, sore throat, hoarseness, fever; and
 CC abnormal chest sounds on auscultation). C. pneumoniae sequence
 CC is also used as vaccines for immunising humans against diseases
 CC caused by C. pneumoniae.
 XX

SQ Sequence 583 AA;

Query Match 81.8%; Score 1850; DB 21; Length 583;
 Best Local Similarity 98.7%; Pred. No. 1.4e-116;
 Matches 381; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 69 MSLADKILGIASSNSSSTSRSDVDSTTATAPPPPTSDVKYKTAQAYDTIFTSTSLA 128
 Db 1 msladkigiasnssstsrSadvdsttatapppttdfdyktqagaydtiftstsla 60
 QY 129 DIQAALVSLQDVTNIDTAATDEETAIAAEWETKNADAIVKGAQITELAKYASDNQAIL 188
 Db 61 diqaalvslqdvtnikdtaateetaiaaewetknadvkvgaqitelakyasdnqail 120
 QY 189 DSLGKLTSLFLLQVANNKAAELLKEMQDNPVPGKTPAIAQSLVDQTDATATQ 248
 Db 121 dslgkltslfllqavannnkaaellkemqdnvpvpgktpaiaqslvdqtdatq 180
 QY 249 IEKDGNAGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTOIAEAKKFPDPSPI 308
 Db 181 iekdgnairdayfagqnasgavenaksnnsisnidsakaaIataktqiaeaqkfpdspI 240
 QY 309 LOEAQOMVIOAEKDLKNIKPADGSDVPNPGTTVGGSKQOGSSIGSIRVSMLLDAENETA 368
 Db 241 lqeaqmviqaeKdlknikpadgsdvpnpgtttvggskqgssigsirvsmlldaeneta 300
 QY 369 SILMSGFROMIHMFNTENPDSSQAQELAAQARAANKAGDDSAALADAQKALEAALGK 428
 Db 301 silmsgfrmiHmfntenpdsgaaqelaaqaraaKaagddsaaladaqkaLeaalgk 360
 QY 429 AGQOQILNALGQTASAAVSVAGVLP 454
 Db 361 agqqgilnalqiasaavsvagvpp 386

RESULT 6

AAV37571
 ID AAV37571 standard; Protein; 350 AA.
 XX
 AC AAV37571;
 XX
 XX
 DT 07-OCT-1999 (first entry)
 XX
 DE Chlamydia trachomatis surface exposed protein.

XX
 XX
 KW Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
 KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
 KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
 KW Bartholinitis; pneumopathy; venereal lymphogranulomatosis.
 XX
 OS Chlamydia trachomatis.
 XX
 XX WO9928475-A2.
 XX
 PD 10-JUN-1999.
 XX

Db 1123 litakadagviaaklaideikledkqaeakaidastmt-----neekaiakkal 1173
Qy 119 DTFTSLADIQAALVSLQDA---VTN- IKDTAATDEETAIAAEWETKNADAIKVGAGI 174
Db 1174 qdv-----vdkgkaeadaarvatneiheattekakaealageksldtckearda 1225
Qy 175 TELAKYASDQAILDSLGK--LTSFDLLQTLALLOSVAANNKAAELLKEMODNPVVPGRK-- 230
Db 1226 velak---dke-----lgkeairteeaatkiveklaedtr-----kaiednplnsdedk 1272
Qy 231 -----TPATAQSLV---DOTDATATOIEKDGNAIGDAYFAGQNASGAVENAKSNNSIS 280
Db 1273 qaeikltdavaktiatrmdnadrtdkrtgeak-aqalad-----lekaketqkia 1320
Qy 281 NIDSAKAI-----AT---AKTOIAE-----AOKKFPDPSILOQAEQWVIOAE 320
Db 1321 d-----kaaidrltlivkgeleatkqdkatkiaakdaaaakaiaasnpnltdaekktfda 1376
Qy 321 KDLKNIKPADGSDVPNPOTTTVGGSKQGGSSIGSIRVSM-----LDAENETASILMSGPRQ 377
Db 1377 vdaevakandaisaatspadv--qkeedagvaataedvldaaqdknkiakdaaaa-ke 1433
Qy 378 MIHMENTENPDSSQAAQQL---AAQARAAGAAGDDSAALADAKALEAALGKAGQQQG 434
Db 1434 al-----gsnpnltdaekktftdavaevakandaisaatspadvqkeeda--gvaataed 1487
Qy 435 ILNALGQIA 443
Db 1488 vldaakqda 1496

RESULT 10

AAy37572
ID AAY37572 standard; Protein; 331 AA.

XX AAY37572;

XX 07-OCT-1999 (first entry)

XX Chlamydia trachomatis cellular envelope protein.

XX Vaccine; eye disease; conventional trachoma; nonendemic trachoma;
KW paratrachoma; inclusion conjunctivitis; genital disease; perihepatitis;
KW nongonococcal urethritis; epididymitis; cervicitis; salpingitis;
KW bartholinitis; pneumopathy; venereal lymphogranulomatosis.

XX Chlamydia trachomatis.

XX WO9928475-A2.

XX 10-JUN-1999.

XX 27-NOV-1998; 98WO-IB01939.

XX 04-NOV-1998; 98US-0107077.

XX 28-NOV-1997; 97FR-0015041.

XX 17-DEC-1997; 97FR-0016034.

XX (GEST) GENSET.

XX Griffais R;

XX WPI; 1999-371125/31.

XX Genome sequence of Chlamydia trachomatis

XX Disclosure; Page 1227-1228; 1755pp; English.

XX AAY36754-Y37949 are encoded by open reading frames (ORFs) of the genome
CC of Chlamydia trachomatis (see AA201425). The polypeptides can be used as
CC vaccines against Chlamydia trachomatis. Antisense and ribozyme sequences
CC can also be used to control growth of the microorganism. Chlamydia
CC trachomatis is responsible for a large number of diseases, e.g. eye

CC diseases such as conventional trachoma, nonendemic trachoma,
CC paratrachoma, and inclusion conjunctivitis; genital diseases such as
CC nongonococcal urethritis, epididymitis, cervicitis, salpingitis,
CC perihepatitis, bartholinitis; pneumopathy in breast feeding infants;
CC and venereal lymphogranulomatosis. The polypeptides of the invention
CC may be of use in treating these diseases.

XX Sequence 331 AA;

Query Match 7.3%; Score 164.5; DB 20; Length 331;
Best Local Similarity 38.5%; Pred. No. 0.0017;
Matches 45; Conservative 12; Mismatches 53; Indels 7; Gaps 3;

Qy 338 GTTVGGSKQGGSSIGSIRVSMLLDDAENETASILMSGFROMIHMENTENPDSSQAAQQLA 397

Db 17 gsaagalk---ssnsgriisllldvdnemaaiamqgfisnieqfnnvmpatakelqame 73

Qy 398 AQARAAGDDSAALADAKALEAALGKAGQQQGGILNALGQIASAAVVSAG 451

Db 74 aqltamsqqlvgadgelpaeiqaikdalaqal-kpsadglatamgqvfaaakvvg 129

RESULT 11

AAy81609

ID AAY81609 standard; Protein; 1237 AA.

XX AAY81609;

XX 24-MAY-2000 (first entry)

XX Streptococcus pneumoniae type 4 protein sequence #109.

XX Streptococcus pneumoniae; vaccine; screening; protein antigen;
KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.

XX Streptococcus pneumoniae.

XX WO200005737-A2.

XX 10-FEB-2000.

XX 27-JUL-1999; 99WO-GB02451.

XX 27-JUL-1998; 98GB-0016337.

XX 19-MAR-1999; 99US-0125164.

XX (MICR-) MICROBIAL TECHNIQS LTD.

XX Gilbert CFG, Hansbro PM;

XX WPI; 2000-195300/17.

XX New Streptococcal protein, useful as a vaccine, for diagnosis of
XX pneumococcal diseases and for screening agents capable of antagonizing
XX or inhibiting expression of the protein

XX Claim 1: Page 83; 108pp; English.

XX AAY81501 to AAY81679 represent specifically claimed protein sequences
CC isolated from Streptococcus pneumoniae. AAY81501 to AAY81590 represent
CC specifically claimed nucleotide sequences isolated from S. pneumoniae.
CC The sequences have antibacterial and antiinflammatory properties.
CC The protein sequences, and fragments of them, are useful as immunogens
CC and/or antigens. The nucleotide sequences can be used in vaccines and in
CC diagnostic assays. The proteins and nucleotides can be useful for the
CC detection and diagnosis of S. pneumoniae. The protein sequences are also
CC useful for screening an agent capable of antagonizing, inhibiting or
CC interfering with the function or expression of the proteins in which the
CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
CC and meningitis. AAY81591 to AAY81614 represent primers used in the
CC exemplification of the present invention.

```

XX SQ Sequence 1237 AA;
Query Match 6.8%; Score 153.5; DB 21; Length 1237;
Best Local Similarity 19.9%; Pred. No. 0.051;
Matches 84; Conservative 84; Mismatches 205; Indels 49; Gaps 9;

QY 13 TERTPPADLSAQGLEASANKSAEQAQRIAGAEKPKESKTDTSVERWILSAVNALMSLA 72
Db 467 tsasasasasasasasasasasasasasasasasasasasasasasasasasas 526
QY 73 DKLGTAGSSSSSTSRGADVDSSTTATPTPTPTPTPTPTPTPTPTPTPTPTPTPT 132
Db 527 ---gsaststasasasasasasasasasasasasasasasasasasasasas 577
QY 133 ALVSLQDAVNIKDTATDETAIAEWEKTKNDAIKVGAQITELAKYASDQAILDSL 192
Db 578 stasasasasasasasasasasasasasasasasasasasasasasasasas 634
QY 193 KLTSLDLOTALLOSVAANNKAAELLKEMQDNPVPGKTPAIAQSLVDQTDATATQ 252
Db 635 asasasasasasasasasasasasasasasasasasasasasasasasas 686
QY 253 CNAIGDAYFAGQNASGAVENAKSNNSISNIDSAK-AAATATAKTAQIAEAKKFPD 311
Db 687 astasasasasasasasasasasasasasasasasasasasasasasasas 741
QY 312 AEQWVIAEKDLKNIKPADGSDVPNPGTTVGGSKQGGSSI-GSTRVSMLLDDAEN 370
Db 742 -----sastsasasasasasasasasasasasasasasasasasasas 785
QY 371 LMSGFRQMIHMFNTENPDPSQAQOELAAQARAAGDSDSAAALADAQAALCKAG 430
Db 786 sas-----esastsasasasasasasasasasasasasasasasasas 839
QY 431 QQ 432
Db 840 rq 841

RESULT 12
AA091443
ID AA091443 standard; Protein; 600 AA.
AC AA091443;
XX
XX
XX
XX 26-SEP-2001 (first entry)
DE C glutamic protein fragment SEQ ID NO: 5197.
KW Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
XX organic acid synthesis.
XX
XX Corynebacterium glutamicum.
OS EP1108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
XX
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW ) KYOWA HAKKO KOGYO KK.
PA
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI: 2001-376931/40.
DR N-PSDB; AAH66662.

```

```

XX
PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
XX Claim 17; SEQ ID NO: 5197; 246pp + Sequence Listing; English.
PS
XX
XX The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX Sequence 600 AA;
Query Match 6.8%; Score 153; DB 22; Length 600;
Best Local Similarity 21.6%; Pred. No. 0.022;
Matches 96; Conservative 66; Mismatches 187; Indels 96; Gaps 14;

QY 18 PADLSAQGLEASANKSAEQAQRIAGAEKPKESKTSVE-RWILSAVN-ALMSLADKL 75
Db 24 pddaaiaqaeevnsagdgevarlags-lsstdaeinrvelmgalreevnxslvdhda 82
QY 76 GIASSNSSSTSRGADVDSSTTATPTPTPTPTPTPTPTPTPTPTPTPTPTPTPT 135
Db 83 aiae-----qarqdaaakddldsqae 107
QY 136 SLODAVTNIKDTATDETAIAEWEKTKNDAIKVGAQITELAKYASDQAILDSLGLT 195
Db 108 aaqerldeisraayrqnqgtskigisngnusedaldrqylrtsaekqqaaveeldrl 167
QY 196 SFDLLQLOTALLOSVAANNKAAELLKEMQDNPVPGKTPAIAQSLVDQTDATATQ 255
Db 168 t-----enankesvlrq-----arivaegreaavekvqtea 200
QY 256 IGDAYFAGQNASGAVENAKSNNSISNIDSAKAAIATAKTAQIAEAKKFPDPSILQEA 315
Db 201 ---aiaanseqlnvltnrs-tlvaqrdgaernlalaiaaraqadnlqggraeeyefqae 256
QY 316 VIOAEKD----LKNIKPADGSDVPNPGTTVGGSKQGGSSIGSTRVSMLLDDAENETAS 371
Db 257 riqaeaeaaqaeeekrradea-aaqaqaeeaaqqaeeaaqqaeeaaqqaeeaaqqa 314
QY 372 MSGFRQMIHMFNTENPDPSQAQOELAAQARAAGDSDSAAALADAQAALCKALEA 425
Db 315 -----aetqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqaqa 362
QY 426 LGKAGQOQOILNALGQIASAAVUSA 450
Db 363 -aqalreqaltaa--siaaaaliaa 384

RESULT 13
AAW02098
ID AAW02098 standard; Protein; 1561 AA.
XX
XX AAW02098;
XX
XX 27-OCT-1996 (first entry)
XX S. mutans antigen I/II.
XX Caries; antigen I/II; epitope; tooth decay; vaccine; immunisation.
KW

```

OS Streptococcus mutans.
 XX Key Location/Qualifiers
 FH Misc-difference 618..650
 FT /note= "amino acids 618-650 differ from the
 FT residues deduced from the nucleic acid
 FT sequence."
 XX
 XX W09623886-A1.
 XX
 XX 08-AUG-1996.
 PD
 XX 31-JAN-1996; 96WO-GB00207.
 XX
 XX 31-JAN-1995; 95GB-0001826.
 PR
 XX (UNME-) UNITED MEDICAL & DENTAL SCHOOLS GUYS.
 PA
 XX Kelly C, Lehner T;
 PI
 XX WPI: 1996-371434/37.
 DR N-PSDB; AAT36122.
 DR
 XX Nucleic acid encoding polypeptide for prevention or treatment of
 PT dental caries - which stimulates T or B cell response, and/or
 PT adheres to tooth in competition with Streptococcus mutans antigen
 PT I/II
 PT
 XX Disclosure; Page 45-46; 63pp; English.
 PS
 XX Streptococcus mutans antigen I/II (AAW02098) is a 185 kDa cell surface
 CC protein at least partly responsible for S. mutans adhesion to teeth.
 CC The I/II antigen includes a series of overlapping T-cell, B-cell and
 CC adhesion epitopes. Fragments (see also AAT36111-21) of the I/II
 CC antigen gene (see also AAT36122) can be used to produce recombinant
 CC polypeptides (AAW02087-97) carrying such epitopes for use in vaccines
 CC for immunisation against dental caries.
 XX
 XX Sequence 1561 AA;

Query Match 6.7%; Score 152; DB 17; Length 1561;
 Best Local Similarity 22.3%; Pred. No. 0.087;
 Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;

QY 48 KESKTDVSRWILRSVAVNALMADKLGIASSN-SSSSTSRSDVDSTATATPTPPPT 106
 DB 2 kvkkygfrkskistlcvlgvtaavsvagkvfadetttsdvtkvvgvtgtnpat 61
 QY 107 -----SDDYKTAQATYDIFTST-----SLADTQAAALVS-LQDAVTNIKDT 147
 DB 62 nlpeaagsaskqaedsqtklerqmvhtievptktdidqaakdaksagvnnvqgdvkn-kgt 120
 QY 148 AATDEETATAAEWE-----TKNADAIK-----VGAQITELAKYASDNOAILDSLGK 193
 DB 121 vtace-avqketelkedytkqaedkktdtdykdsvaheaeavakikakngatkeqyqk 179
 QY 194 LFSFOLL-OTALLQSVANNKAAEL-----LKEMQDNVPVPOKTPAIAQSLVDQDATAT 247
 DB 180 ----dmvnhkaeaverinaanaaskatayeaqlaqyqadlaavqktnaanaqasy-qkalaay 234
 QY 248 QIE----KQGNAGIDAYFAGQNAAGAVENAKSNNSISNIDSAAKAIATAK-----TQIAEA 299
 DB 235 qaelkrvgeanaakaay---dtavaannakntelaanaeeirkrnaktaeayeklaqy 291
 QY 300 Q---KKFPDPIQAE-QMVIQA-EKDLKNIKPAD-----GSDVPNPGTTVG 342
 DB 292 qaelkrvgeanaaneadyqaktayqtelarvqkanadakaayaavaannaknaaltae 351
 QY 343 GS-----KQGSISIGIRVSMMLDIAENEFASILMSGFROMIHWFNTE 385
 DB 352 ntaikqrnenakatyeaalkqyeadlaavkvkanaaneadyqak---ltayqtelarvqka 408

QY 386 NFDSSAAQOELAAQAAKAA--GDSAAAL-ADAQKALEALCK 428
 DB 409 nadakaayaavaannaaltaentaikkrnadakadyeaklak 454
 RESULT 14
 AAR84568
 ID AAR84568 standard; Protein; 643 AA.
 XX AAR84568;
 AC
 XX 09-MAY-1996 (first entry)
 DT
 XX Trypanosoma cruzi TCR27 polypeptide, Ag15.
 DE
 XX TCR27; Chagas disease; repeat unit; diagnosis; blood screening;
 KW recombinant; fusion protein; glutathione-S-transferase.
 XX
 XX Trypanosoma cruzi.
 OS
 XX Key Location/Qualifiers
 FH 329..552
 FT /label= repeat_region
 FT /note= "16 of 69 repeat units of 14 amino acids"
 FT
 PN W09525797-A1.
 XX
 XX 28-SEP-1995.
 PD
 XX 20-MAR-1995; 95WO-US03191.
 PF
 XX 24-MAR-1994; 94US-0216894.
 PR
 XX (KIRC/) KIRCHHOFF L V.
 PA (ORSU/) OFSU K.
 PA
 XX Kirchhoff LV, Otsu K;
 PI
 XX WPI: 1995-344618/44.
 DR N-PSDB; AAT05332.
 DR
 XX New polypeptide(s) from the TCR27 protein of Trypanosoma cruzi - as
 PT immunoassay reagent for specific diagnosis of Chagas disease, also
 PT related nucleic acid and transformed cells
 PT
 XX Disclosure; Page 40-41; 68pp; English.
 PS
 XX AAR84565-R84569 are polypeptides of the TCR27 protein of T. cruzi
 CC The proteins are all fusion products with glutathione-S-transferase
 CC (GST) and some contain a linker sequence. The TCR27 protein comprises
 CC a 95 amino acid (aa) N-terminal region; 69 repeats of a highly
 CC conserved 14 aa sequence and a 68 aa C-terminal region. This sequence
 CC encodes the GST sequence, the A944 polypeptide contg. 16 of the 69
 CC repeat units and also contains the amino and carboxy terminal
 CC peptides of TCR27. The TCR27 polypeptides of the invention are useful
 CC for the diagnosis of Chagas disease (American Trypanosomiasis), they
 CC are capable of detecting anti-T. cruzi antibodies; or for blood
 CC screening. The TCR27 protein has epitopes to which most T. cruzi
 CC infected individuals have antibodies. The TCR27 polypeptides will not
 CC react with serum from patients with leishmaniasis, schistosomiasis,
 CC or autoimmune disease and are hence less likely to cause false
 CC positives in diagnosis.
 XX
 XX Sequence 643 AA;

Query Match 6.7%; Score 150.5; DB 16; Length 643;
 Best Local Similarity 21.5%; Pred. No. 0.035;
 Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps 17;

QY 7 PGPIDETPTPADISAQGLEASAAKSAQRIAGAEAKPKESKTSVSRVILRSVAVN 66

Db 216 ppsdvlprgsqll--qgaennitnskkmkmlrekvkkaekeldainratkleeern 273
 Qy 67 ALMSLADKLGASSNSSSTSRSDVDSTTATPTPPPTSDDYKTKQAQTAAYDTFTSTS 126
 Db 274 qaykaahk---aeekaktfgrlltfseninlkkrp----- 307
 Qy 127 LADIAQALVSLQDAVN-----IKDTAATDE-----ETAIAAEWETKNADAIVGAQAI 174
 Db 308 -----ndavsnrdkkknsetaktdevekraaeaaakavetekqra-----aea 350
 Qy 175 TELAKYASDNQALDLSGLKTSFDLLQTALLQSVANNNAKAEELKEMQDNVPPVPGKTPAI 234
 Db 351 lkvae-----akrkkaaeakavetekqraaeatk 381
 Qy 235 A-----QSLVDQTDATATQIEKDNAGDAYFAGQNASCAVENAKSN-----NSISNID 283
 Db 382 aeekqkaaeakavetekqraaeatkvaeeakqraaeamkvaeeakqkaaeatkvaee 441
 Qy 284 SAKAAATATAKTQIAEAQR-KFPDPSILOEAPQM-----VIOAEKDLKNIKPADGSDVP 335
 Db 442 kqkaaea---tkvaeaekkaaeatkvaeeakqkaaeatkvaeeakq-----kaaeatkva 494
 Qy 336 NPETTVGSGKQSGSIGIRVSMLLDDAENETASILMSGFROMIHMFTENPDSQAAQOE 395
 Db 495 e-----aekqkaaeatkva-----eaekqkaaeatk-----vaeaekqkaaeat 533
 Qy 396 LAAQARAAGAAGDDSAALADAKALEAALGKAGQOQIL 436
 Db 534 kvaeaekkaa---eatkvaeeakqkaaeakamesqkgrfl 572

RESULT 15
 ID AAB19849
 XX AAB19849 standard; Protein; 729 AA.
 AC AAB19849;
 XX 05-MAR-2001 (first entry)
 XX Mycobacterium tuberculosis protein MTB8.
 XX MTB8; tuberculosis; BCG; vaccine; infection; diagnosis.
 OS Mycobacterium tuberculosis.
 PN WO2000066157-A1.
 XX 09-NOV-2000.
 XX 04-MAY-2000; 2000WO-US12257.
 PR 04-MAY-1999; 99US-0132505.
 PA (PUBL-) PUBLIC HEALTH RES INST NEW YORK.
 PI Gennaro ML;
 DR WPI; 2001-007153/01.
 DR N-PSDB; AAA89042.
 XX Novel polypeptide encoded by open reading frames present in
 PT Mycobacterium tuberculosis genome and not by the BCG strain of M.
 PT bovis, useful as vaccine and for diagnosing tuberculosis infection
 XX Claim 11; Fig 1; 35pp; English.
 CC The present sequence is that of the Mycobacterium tuberculosis
 CC MTB8 protein. This is 1 of 8 proteins, i.e. MTB8-1 (see
 CC AAB19842-49), encoded by 8 open reading frames (see AAA89035-42)
 CC identified as being present in the genome of M. tuberculosis but
 CC absent from the genome of the BCG strain of Mycobacterium bovis.
 CC MTB8-1 represent reagents that are useful in discriminating between
 CC M. tuberculosis and BCG and, in particular, for diagnostic methods

CC which discriminate between exposure of a subject to M. tuberculosis
 CC and vaccination with BCG. The invention features these MTB8
 CC polypeptides, functional fragments of them, DNA encoding them,
 CC vectors, transformed cells, and diagnostic, therapeutic, and
 CC prophylactic (vaccine) methods, including genetic vaccination
 CC methods.
 XX Sequence 729 AA;
 SQ
 Query Match 6.5%; Score 148; DB 22; Length 729;
 Best Local Similarity 23.2%; Pred. No. 0.06;
 Matches 103; Conservative 46; Mismatches 181; Indels 114; Gaps 17;
 Qy 6 GPGPIDETER--TPPADLSAQGLEASAAKSAQAQRTAGAEAKPKESKTSVERWSILRS 63
 Db 351 gpahadesaasvtpaaasgvpgraaapsgta---vgaga-----rs 391
 Qy 64 AVNALMSLADKLGIASSNSSSTSRSDVDSTTA---TAPTPPPTSD--DYKTKQAQTA 117
 Db 392 svgtaaasgagshaatgrapvatsdkaaapstraasartapparpsthdidkpdresa 451
 Qy 118 YDTFTSTSLADIQAALVSLQDAVTNKTDAATDEETAIAAEWETKNADAIVGAQITEL 177
 Db 452 ddg--tpvsmipvsaa-----raardaataaasargrgrgdalrria-a 495
 Qy 178 AKVASDNOAILDLSGLKTSFDLLQTALLQSVANNNAKAEELKEMQDNVPPVPGKTPAIQS 237
 Db 496 alnasdna-----gdygffwi----- 512
 Qy 238 LVDQTDATATQIEKDG-----NAIGDAYFA-GQNASGAVENAKSNNSISNIDSAKAIAT 291
 Db 513 -----tavtdgsvivvansyglayipdgmelpnkvyilasadhaip-vdei-arcat 561
 Qy 292 AKTOIAEAQKFFPD---SPILQEAQMVIOAEKDLKNIKPADGSDVPNPGTTVGSQKQ- 347
 Db 562 ypvlavqawaafhdmtravigttaeqlassdpgvakivlepd--dipesgkmtgrslev 619
 Qy 348 -----GSSIGSIRVSMLLDDAENETASILMSGFROMIHMFTENPDSQAAQOEALAAARA 402
 Db 620 vdpasaaqladttdqrlldllppapvdpnpgderhmlwfeilmkpmststatgreaahlra 679
 Qy 403 AKAGDDSAALADAKALEAAL 426
 Db 680 frayaahsgeialhqahtatdaav 703

Search completed: February 7, 2002, 21:34:38
 Job time: 20757 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: February 7, 2002, 21:36:19 ; Search time 85.91 Seconds
(without alignments)
121.278 Million cell updates/sec

Title: US-09-391-606-9

Perfect score: 2261
Sequence: 1 MWNPIGPGRIDETERRPPAD.....SAAVVSAGVPLQGVLMIRA 463

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 212252 seqs, 22503292 residues

Total number of hits satisfying chosen parameters: 212252

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCtUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/backfiles1.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	152	6.7	1561	3 US-08-894-017-23	Sequence 23, Appl
2	150.5	6.7	643	2 US-08-216-894-8	Sequence 8, Appl
3	150.5	6.7	643	4 US-09-115-746-8	Sequence 8, Appl
4	146.5	6.5	1565	6 5352450-2	Patent No. 5352450
5	143	6.3	1566	2 US-08-687-956A-23	Sequence 23, Appl
6	138.5	6.1	564	2 US-08-216-894-2	Sequence 2, Appl
7	138.5	6.1	564	4 US-09-115-746-2	Sequence 2, Appl
8	138	6.1	2842	1 US-07-741-940-7	Sequence 7, Appl
9	138	6.1	2842	1 US-08-289-548A-7	Sequence 7, Appl
10	138	6.1	2842	1 US-08-452-654-7	Sequence 7, Appl
11	138	6.1	2843	1 US-07-741-940-2	Sequence 7, Appl
12	138	6.1	2843	1 US-08-289-548A-2	Sequence 2, Appl
13	138	6.1	2843	1 US-08-452-654-2	Sequence 2, Appl
14	138	6.1	2843	1 US-08-452-655B-2	Sequence 2, Appl
15	138	6.1	2843	1 US-08-452-655B-7	Sequence 7, Appl
16	138	6.1	2843	2 US-08-370-235A-2	Sequence 2, Appl
17	138	6.1	2843	3 US-08-450-582-2	Sequence 2, Appl
18	138	6.1	2843	3 US-08-450-582-7	Sequence 7, Appl
19	138	6.1	2973	2 US-08-821-355A-7	Sequence 7, Appl
20	138	6.1	2973	2 US-09-003-687A-7	Sequence 7, Appl
21	138	6.0	2285	4 US-09-136-605-7	Sequence 7, Appl
22	136.5	6.0	2285	4 US-09-308-375-2	Sequence 2, Appl
23	135.5	6.0	639	1 US-08-466-390-2	Sequence 2, Appl
24	135.5	6.0	639	1 US-08-470-950-2	Sequence 2, Appl
25	135.5	6.0	639	1 US-08-467-781-2	Sequence 2, Appl
26	135.5	6.0	639	1 US-08-195-487-2	Sequence 2, Appl
27	135.5	6.0	639	2 US-08-483-924-2	Sequence 2, Appl

28	135.5	6.0	639	5 PCT-US93-06160-2	Sequence 2, Appl
29	131.5	5.8	593	2 US-08-591-079-8	Sequence 8, Appl
30	131.5	5.8	593	2 US-08-591-079-10	Sequence 10, Appl
31	127.5	5.6	635	4 US-08-931-608A-5	Sequence 5, Appl
32	126	5.6	466	4 US-08-235-836C-107	Sequence 107, App
33	123.5	5.5	461	1 US-08-186-222-2	Sequence 2, Appl
34	123.5	5.5	619	2 US-08-372-652-5	Sequence 5, Appl
35	123.5	5.5	619	3 PCT-US95-16311-5	Sequence 2, Appl
36	123.5	5.5	933	3 US-08-293-728-2	Sequence 2, Appl
37	123.5	5.5	933	4 US-09-421-868-2	Sequence 11, Appl
38	123.5	5.5	2756	1 US-08-375-709-11	Sequence 11, Appl
39	123.5	5.5	2756	1 US-08-752-929-11	Sequence 11, Appl
40	123.5	5.5	2756	4 US-09-050-793-7	Sequence 7, Appl
41	123	5.4	1912	1 US-08-409-995-4	Sequence 4, Appl
42	123	5.4	1912	3 US-08-685-467-4	Sequence 4, Appl
43	123	5.4	2353	4 US-09-377-155-33	Sequence 33, Appl
44	123	5.4	2353	4 US-08-913-942-4	Sequence 4, Appl
45	122.5	5.4	1196	1 US-08-144-121-4	Sequence 4, Appl

ALIGNMENTS

RESULT 1
US-08-894-017-23
Sequence 23, Application US/08894017
Patent No. 6024958
GENERAL INFORMATION:
APPLICANT: Lehner, Thomas
TITLE OF INVENTION: POLYPEPTIDE FRAGMENTS CAPABLE
OF COMPETITION WITH STREPTOCOCCUS MUTANS ANTIGEN I/II
NUMBER OF SEQUENCES: 27
CORRESPONDENCE ADDRESS:
ADDRESSEE: MORRISON & FOERSTER
STREET: 2000 Pennsylvania Avenue, NW, suite 5500
CITY: Washington
STATE: DC
COUNTRY: USA
ZIP: 20006-1888
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows
SOFTWARE: FASTSEQ for Windows Version 2.0b
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/894,017
FILING DATE: 31-JUL-1997
CLASSIFICATION: 536
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/GB96/00207
FILING DATE: 31-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Murashige, Kate H
REGISTRATION NUMBER: 29,959
REFERENCE/DOCKET NUMBER: 25150-20067.00
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-887-1500
TELEFAX: 202-822-0168
TELEX:
INFORMATION FOR SEQ ID NO: 23:
SEQUENCE CHARACTERISTICS:
LENGTH: 1561 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-894-017-23
Query Match 6.7%; Score 152; DB 3; Length 1561;
Best Local Similarity 22.3%; Pred. No. 0.004;
Matches 104; Conservative 82; Mismatches 182; Indels 98; Gaps 23;

QY	48	KESKDVSERNISLSBSANMLMSLADKIGLSSN-SSSSIRSDADYSTATAATPEPPT	106
Dd	2	KVKKTYGRKRSKISITLLOGAVLGTVAANVSAGQVPEADETTTTSDVYTPKVVGTQGNPAT	61
QY	107	-----SDDYKTQAOATAYDTIFTST-----SLADIQAALVS-LQDAVNTIKDT	147
Dd	62	NLPEAGSASQAOESOTKTEROWHVTIEVPEKTDLDQAARKDAKSAGVNVQDADVFN-KGT	120
QY	148	AATDEETIAAEWE-----TKNAAIK-----YGAOTTELAKASDNOALDSLGK	193
Dd	121	VKTAE-VAQETEKEDYTKQADKIKTTDOYSDVAHAEVAKIKRANQAKREYQK	179
QY	194	LTSFDLL-OTALLQSVANNKRAEL-----LKEQMDNPVVPYGPKTPAIQSLSVDTQATAT	247
Dd	180	----DMVAHKAEVEINANANASTAYEAKIQAQYQADLAVQKINANQMSY-QKALAAV	234
QY	248	QIE---KDGNAIGDAYFAGONASGAVENAKSNSISINTSAKAIAITAK---TOJIAEA	299
Dd	235	QAEIKRVOEANAANKAAV---DTVVAANMANTEIAAANEIRKRNATIAAEVETTKIQAQY	291
QY	300	O---KKPFDSPILQEA-OWVIOA-EKDLNKRQAD-----GSDVPRPGTYS	342
Dd	292	QAEIKRVOEANAANEAQAKLTLYOTELAVQANADAKAAYEAVAANNAKAAALTAE	351
QY	343	GS-----KOOGSSIGSIRVSMLLDDAENETASILMSGFROMTMENTE	385
Dd	352	NTAIKORNENNAKATYEALQYUEDLAAVKKANANAEADYQAK---LTAVQTELAVQKA	408
QY	386	NPDQQAQOELAAQARAKAA--GDDSAAL-ADAQKALEALGK	428
Dd	409	NADRAAYEAVAANNAANALTLMENTAIKRNADADADAEYETAK	454

RESULT 2
 ; US-08-216-894-8
 ; Sequence 8, Application US/08216894
 ; Patent No. 5876734
 ; GENERAL INFORMATION:
 ; APPLICANT: Kirchhoff, Louis V.
 ; APPLICANT: Otsu, Keiko
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
 ; TITLE OF INVENTION: WITH TRIPANOSOMA CRUZI
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/216,894
 ; FILING DATE: 24-MAR-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 85326/102/DRL0
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; TELEX: 904136
 ; INFORMATION FOR SEQ ID NO: 8:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 643 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-216-894-8

```

Query Match=          6.7%; Score 150.5; DB 2; Length 643;
Best Local Similarity 21.5%; Pval No. 0.0015;
Matches 99; Conservative 58; Mismatches 169; Indels 135; Gaps

QY      7  PGPIDERTPPADLSAAGLEASAAKSAEAOIAGAEPKESKTDSVERMSILRSVN 66
Db      216  PKSPDLVPRGSPSQL--QAENNTNISKEMTKLREYKAKEKELDINRATKLEERN 273
QY      67  ALMSLADKLGIASSNSSSSSTSRSDVDSTTATPTATPPPSPDDXKYTAQTAYDTIFSTS 126
Db      274  QAYKAAH---AEBKATFORLTTFSEENINKRP-----307
QY      127  LADIQAALVSLQDAVTN----IKDTAATDE--S---ETAIAEWETKNADAIKVGAOI 174
Db      308  -----NDAVSNNDKKNNSETAKTDVEVEKQRAEAKAKEATVKORA-----AEA 350
QY      175  TELAKYASNDAQIILDSIGKLTSPDLLQTALLQSVAANNKKAELLKEQNDPVPVGKTPAI 234
Db      351  TKVAE-----AEKRKAEEAAKAAVETEKOORAAEATKV 381
QY      235  A-----OSLVQDTQATPQIEKDGNAILGDVAFAGONASGAVENAKSN-----NSTSNID 283
Db      382  AEAEKOKAAEAKAAVETEKOPAAEATKYAEAEKORPAEAMKVMAEAEKOKAAEATKYAEAE 441
QY      284  SAKAAIATAKTOIAEAK-KFPDSPIIQAEOM-----VIOAEKDLKNIKIPADGSVP 335
Db      442  KORAEEA--TKVAEAEKOKAAEATKYAEAEKOKAAEATKYAEAEKQ-----KAAEATKYA 494
QY      336  NPGITVGGSKQOQSGISIRISMLLDOAENHTASILMSGFQMTHMNTENPNPSQAAQE 395
Db      495  E-----AEKKAEEATKYA-----FAEKOKAAEATK-----VAEEKOKAAEAT 533
```

RESULT 3
 US-09-115-746-8
 Sequence 8, Application US/09115746
 Patent No. 6228601
 GENERAL INFORMATION:
 APPLICANT: Kirchhoff, Louis V.
 APPLICANT: Otsu, Keiko
 TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
 TITLE OF INVENTION: WITH TRIPANOSOMA CRUZI
 NUMBER OF SEQUENCES: 10
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: Foley & Lardner
 STREET: 3000 K Street, N.W., Suite 500
 CITY: Washington, D.C.
 COUNTRY: USA
 ZIP: 20007-5109
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/115,746
 FILING DATE:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/216,894
 FILING DATE:
 ATTORNEY/AGENT INFORMATION:
 NAME: BENT, Stephen A.
 REGISTRATION NUMBER: 29,768
 REFERENCE/DOCKET NUMBER: 85326/102/DRLO
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (202)672-5300
 TELEFAX: (202)672-5399

TELEX: 904136
 INFORMATION FOR SEQ ID NO: 8:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 643 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-115-746-8

Query Match 6.7%; Score 150.5; DB 4; Length 543;
 Best Local Similarity 21.5%; Pred. No. 0.0015;
 Matches 99; Conservative 56; Mismatches 169; Indels 135; Gaps 17;

QY 7 PGPIDEFERPPADLSAGLEASANKSAEAOIAGAEAKPREKSTDSVERMSILRSVN 66
 DB 216 PPKSDLVPRGSPOL--QQAENNTINSKEMTKLEKVKKAKEKLDAINRATKLEERN 273
 QY 67 ALMSLADLGIASSNSSSTSRADVDSTTATPTPPPTSDDYKTOATVDITFTSTS 126
 DB 274 QATRAAK--AEERKAKTFORLTFESEINLKRP----- 307
 QY 127 LADIAALVSLQDAVTN-----IKDTATDE-----ETAIAEMETKNADAIKVAOI 174
 DB 308 -----NDAVSNDRKKNSSETAKTDEVEKQRAAEAKAVETEKQRA-----AEA 350
 QY 175 TELKAYSNDQAIIDSLGKLTSPDLLQTAALQSVANNKKAELKEMODNPVPEKTPAI 234
 DB 351 TKVAE-----AEKRAAEAKAVETEKQRAAEATKV 381
 QY 235 A-----OSLVDTDAVATQIEKDGNAIGDAYFAGONASGAVENAKSN-----NSISMD 283
 DB 382 AEAEKQKAAEAKEAVEKQRAAEATKVAEAEKQRAAEAKVAEAEATKVAEAE 441
 QY 284 SAKAIAITAKTQIAEAK--KEPDSPILOEADM-----VIOEKDLKNIPADGSDVP 335
 DB 442 KQKAAEA---TKVAEAEKQKAAEATKVAEAEKQKAAEATKVAEAEKQ---KAAEATVA 494
 QY 336 NPGTGVGSKQSGSIGIRVSMILDDAENETASTILMSGFQMIHMTENDPSQAQOE 395
 DB 495 E-----AEKQKAAEATKVA-----EAEKQKAAEATK-----VAEAEKQKAAEAT 533
 QY 396 LAAQRAAKAGDDSAALALADAKALAEALGAKGQOOGIL 436
 DB 534 KVAEAEKQKAA--EATKVAEAEKQKAAEAKAAMESQKORFL 572

RESULT 4
 5352450-2
 Patent No. 5352450
 APPLICANT: KOGA, TOSHIOHKO; OKAHASHI, NOBUO; TAKAHASHI, ICHIRO;
 SHIBUYA, KOJI; OHRA, HIROTAKA
 TITLE OF INVENTION: METHOD FOR PREPARING VACCINE FOR DENTAL
 CAVITIES AND VACCINAL COMPOSITIONS FOR DENTAL CAVITIES USED AS NASAL
 DROP
 NUMBER OF SEQUENCES: 2
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/07/529,602
 FILING DATE: 29-MAY-1990
 SEQ ID NO: 2
 LENGTH: 1565
 5352450-2

Query Match 6.5%; Score 146.5; DB 6; Length 1565;
 Best Local Similarity 21.8%; Pred. No. 0.011;
 Matches 101; Conservative 81; Mismatches 190; Indels 91; Gaps 21;
 QY 48 KESTDSVERMSILRSVANNALMSLADKLGIASSN--SSSSSRADVDVSTAAEATPPPP 105
 DB 2 KVKTTTGRKRSKISKILGAVIGVAAVAVAGQVFADETTTTSVDVTKVVGTGTGMPA 61
 QY 106 T-----SDDYKTOATVDITFTST-----SLADIAALVS-LQDAVTNIKD 146

DB 62 TQLPEAGSASKEABQSOTKLERQWHTIEPKTDLDQAAKAGVNVQDADVN--KG 120
 QY 147 TAATDEETAIAAEWE-----TKNADAIK-----VGAQTTELAKYASDQALDLSLG 192
 DB 121 TKTPEE-AVOKETIEKEDYTKQADIKTTDQTSVDAHAHEAEVAKKAKKQATKEQYE 179
 QY 193 -KLTSPDLLQTAALQSVANNKKA--ELLKEMODNPVPEKTPAIAQSLVDQTDATATQIE 250
 DB 180 KDMAHKAEEVERINANANASKATYAEAKLAQYQADLAAYQKTAANQAAV--QALAAAYQAE 238
 QY 251 ----KDGAIAIGDAYFAGONASGAVENAKSNNSISIDSAKAIATAK-----TQIEAQ-- 300
 DB 239 LKRVQEAANNAKAAV--DTAAVANNANNTETIAAANEIRKRNATYAKAEYETKILQYQAE 295
 QY 301 -KEPDSPILOEAE--QNVIOA--EKDLKNIKPAD-----GSDVPNPPTGVGS- 344
 DB 296 LKRVQEAANNAEADYQALUTAYQTELARVQKANAADAKATYEAANAANKNAALTAENTA 355
 QY 345 -----KQSGSIGIRVSMILDDAENETASTILMSGFQMIHMTENDP 388
 DB 356 IKERNENAKAYEALQYEAALQYEAALQYEAALQYEAALQYEAALQYEAALQYEAALQY 412
 QY 389 SOAAQOELAAQARAKAA--GDSAAAL--ADQKALEALCK 428
 DB 413 AKAAEAAVANNANNAALTAENTAIAIKRNADAKADYEKILAK 455

RESULT 5
 US-08-687-956A-23
 Sequence 23 Application US/0867956A
 Patent No. 5861157
 GENERAL INFORMATION:
 APPLICANT: BURNIE, JAMES P
 APPLICANT: MATTHEWS, RUTH C
 TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF
 TITLE OF INVENTION: INFECTIONS DUE TO STREPTOCOCCI AND ENTEROCOCCI
 NUMBER OF SEQUENCES: 23
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: PILLSBURY, MADISON & SUTRO, LLP
 STREET: 1100 NEW YORK AVENUE, N.W., EAST TOWER, 9TH
 CITY: WASHINGTON
 STATE: DC
 COUNTRY: USA
 ZIP: 20005
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.25
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/687,956A
 FILING DATE: 29-JUL-1996
 CLASSIFICATION: 536
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: GB 9401689.6
 FILING DATE: 28-JAN-1994
 ATTORNEY/AGENT INFORMATION:
 NAME: KOKULIS, PAUL N
 REGISTRATION NUMBER: 16,773
 REFERENCE/DOCKET NUMBER: 50885/222892
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: 202/861-3000
 TELEFAX: 202/822-0944
 INFORMATION FOR SEQ ID NO: 23:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 1566 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: unknown
 MOLECULE TYPE: peptide
 HYPOTHETICAL: NO

ANTI-SENSE: NO
 FRAGMENT TYPE: Internal
 ORIGINAL SOURCE:
 ORGANISM: Streptococcus sobrinus
 US-08-687-956A-23

Query Match 6.38; Score 143; DB 2; Length 1566;
 Best Local Similarity 22.88; Pred. No. 0.019;
 Matches 94; Conservative 67; Mismatches 179; Indels 72; Gaps 17;

```

QY 48 KESTDVSERMSILRSANALMSLADKIGIASNSSSSSTS--RSADVSTATATATPPPP 105
DB 10 KRSISRLGALLGTAL--LASVYGKALAEETSTSTGVNTAVGTETGNPATNLPD 67
QY 106 TSDYKTOQTAY-----DTFTSTLADIQALVSLQDAVTNKDTAIDEETAIA 158
DB 68 KQDPSSQAEFTSQAGKGTGMSVSTSELEDAKSAQEGVTVSQDAFVDKCTVETS 127
QY 159 EWEKKNADAK--VGAQITELAKYASD-----NQALIDSLGKITSFDLQTAALQSYA 209
DB 128 DEANQKETEIKDDYSKQADIQKTEDEYKAAVARNAQETD---RTQENAAKKAQYEDL 184
QY 210 NNNKAEELKEMODNPVVPKTPALQSLVQDTATQIEKQNAIGDAYFAGQNASGA 269
DB 185 AANK-AEVERTNEN-----AQKKADY-EAKLAQYQKDLAAVQQAANDSQAAATA 232
QY 270 VEMAKSNSISNTDSAKAIAITAKTQIAEA-----QKKFDPSPILQ-----EAE 313
DB 233 AKEA-YDEKLARVQAANA--AKKEYEALAAVTNKQIEKAENALIQORNAQAKADYE 288
QY 314 QMVIQAEKDLNIRKADSDVNPNGTIVGSGSQGSSSTGTSIRVSLDDAENETSIIMS 373
DB 289 AKLQYERKDLAAQSGNATNEADYQAKRAVEQELARV-----QAAVAAK---Q 335
QY 374 GFRQIHMFNTENPDQQAQOELAAQARAKAGDSSAAALADDAQKALEAA 425
DB 336 AVEQALANTAKN--AQITAEENALIQORNAQAKA--NVEAKLAQYQKDLAA 383

```

RESULT 6
 US-08-216-894-2
 ; Sequence 2, Application US/08216894
 ; Patent No. 5876734
 ; GENERAL INFORMATION:
 ; APPLICANT: Kirchhoff, Louis V.
 ; APPLICANT: Otsu, Keiko
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
 ; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/216,894
 ; FILING DATE: 24-MAR-1994
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: BENT, Stephen A.
 ; REGISTRATION NUMBER: 29,768
 ; REFERENCE/DOCKET NUMBER: 85326/102/DRLO
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (202)672-5300
 ; TELEFAX: (202)672-5399
 ; INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
 LENGTH: 564 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-08-216-894-2

Query Match 6.1%; Score 138.5; DB 2; Length 564;
 Best Local Similarity 21.4%; Pred. No. 0.01;
 Matches 97; Conservative 54; Mismatches 165; Indels 137; Gaps 17;

```

QY 7 PGPIDETERTPPADLSAGLESAANKSAEQRINGAEKPREKSTDSVERMSILRSANV 66
DB 216 PPKSDLVPRGSPSL--QQAENNITNSKKEMKLEKRYKKAKEKLDINRTKLEBERN 273
QY 67 ALMSLADKIGIASNSSSSSTSRSADVSTTATATPPPTSDYKTOQTAVDITFTSTS 126
DB 274 QAYKAHK--AEKAKTFFORLIIFESFNILKKRP----- 307
QY 127 LADIQALVSLQDAVTN---IKDTATDE-----ETAIAMETKKNADAIYGAOI 174
DB 308 -----NDAVSNRDKKKNSETAKTDEYKORAAEAAKAVETEKORA-----AEA 350
QY 175 TELAKYASQNALIDSLGKITSFDLQTAALQSVANNKKAELLKEMODNPVVPKTPAI 234
DB 351 TKVAE-----AEKRAAQAQAAVETEKORAAEATKV 381
QY 235 A-----QSLVDQDATATQIEKQNAIGDAYFAGQNASGAVENAKSNSISNTDSAKAIA 290
DB 382 AAEKQKAEEAKKAVETEKORAAEATKVAEAEKQAAEAMK-----VAEAEKQKAABA 434
QY 291 TAKTOIAENQK-KFPDPSPLQEAQOM-----VIOAEKDLNIRKADSDVNPNGTIVG 342
DB 435 ---TVVAEAEKQKAEEATKVAEAEKQKAEEATKVAEAEKO---KAAETKVAE----- 481
QY 343 GSKQGSSSTGTSIRVSLDDAENETASILMSGFRQIHMFNTENPDQQAQOELAAQARA 402
DB 482 AEKQKAEEATKVA-----EAEKQKAEEATK-----VAEAEKQKAEEATKVAEAEK 526
QY 403 AKAAGDSSAAAA---LADQKALEALGKAGQ 431
DB 527 OKAAEATKVAEAEKQKAEEATKVAEAEKQKAGE 559

```

RESULT 7
 US-09-115-746-2
 ; Sequence 2, Application US/09115746
 ; Patent No. 6228601
 ; GENERAL INFORMATION:
 ; APPLICANT: Kirchhoff, Louis V.
 ; APPLICANT: Otsu, Keiko
 ; TITLE OF INVENTION: POLYPEPTIDES FOR DIAGNOSING INFECTION
 ; TITLE OF INVENTION: WITH TRYPAOSOMA CRUZI
 ; NUMBER OF SEQUENCES: 10
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Foley & Lardner
 ; STREET: 3000 K Street, N.W., Suite 500
 ; CITY: Washington, D.C.
 ; COUNTRY: USA
 ; ZIP: 20007-5109
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/115,746
 ; FILING DATE:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/216,894
 ; ATTORNEY/AGENT INFORMATION:

NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 85326/102/DRLO
TELEPHONE: (202)672-5300
TELEFAX: (202)672-5399
TELEX: 904136
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 564 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-09-115-746-2

Query Match 6.1%; Score 138.5; DB 4; Length 564;
Best Local Similarity 21.4%; Pred. No. 0.01;
Matches 97; Conservative 54; Mismatches 165; Indels 137; Gaps 17;
QY 7 PGPIDERTPPADLSAOGLEASAAKSAEQAORAGAEAKPKESKTDVSVERWSILRSASN 66
DB 216 PPKSRLVPRGSPQL--QOENNITNSKKEMTKLREKVKAEKELDRINRATKLEERN 273
QY 67 ALMSLADKLGIASSNSSSSRSADVDSTTATPTPPPTSDDYKTOAQATYDTFTSTS 126
DB 274 QAYKAAHK---ABEERAKTFORLITPESINILKKRP----- 307
QY 127 LADIAQALVSLQDAVN-----IKDTAATDE-----ETAIAAEWETKNADAIVCAQI 174
DB 308 -----NDVSNRDKKKKNSSETAKTDEVEKQRAAEAAKAVETEKORA-----AEA 350
QY 175 TELAKYASNDQALDLSGLKLTSPDLLQTLQSVANNKAAELLLKEMODNPVVPKTPAI 234
DB 351 TKVAE-----AEKKAEEAAKAVETEKQRAAEATKV 381
QY 235 A-----QSLVQDTATQIEKDGNAIGDAYFAGONASGAVENAKNSNISIDSAKAAIA 290
DB 382 AEAKQKAAEAAKAVETEKQRAAEATKVAAEKQRAAEAMK-----VAEAEKQKAAEA 434
QY 291 TAKTOIAEAK-KFPDPIQAEOM-----VIOAEKDLKNIKPADGSDVNPNGTVG 342
DB 435 ---TKVAEAEKQRAAEATKVAAEKQKAAEATKVAAEAKO-----KAAEATKVAE- 481
QY 343 GSKQOQSGSIRVMSLLDADENETASILMSGFRQIMHFNENTNPDSQAAQQLAAQARA 402
DB 482 AEKQKAAEATKVA-----EAERKAAEATK-----VAEAEKQKAAEATKVAAEAK 526
QY 403 AKAGDDSAANA-----LADAKALEAALGKAGQ 431
DB 527 QKAAEATKVAAEKQKAAEATKVAAEKQKAGE 559

RESULT 8

US-07-741-940-7

Sequence 7, Application US/07741940

Patent No. 5352775

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie & Beckett

STREET: 1001 G Street, NW
CITY: Washington
STATE: D.C.
COUNTRY: USA
ZIP: 20001-4598
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA: US/07/741,940
APPLICATION NUMBER: US/07/741,940
FILING DATE: 19920109
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Kagan, Sarah A.
REGISTRATION NUMBER: 32,141
REFERENCE/DOCKET NUMBER: 1107.035574
TELEPHONE: 202-508-9100
TELEFAX: 202-508-9299
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 2842 amino acids
TYPE: AMINO ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
IMMEDIATE SOURCE:
CLONE: APC
US-07-741-940-7

Query Match 6.1%; Score 138; DB 1; Length 2842;

Best Local Similarity 20.0%; Pred. No. 0.11;

Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

QY 11 DETERTPPADLSA-OGLEASAAKSAEQAORAGAEAKPK-----SKTDSV----- 55
DB 1658 DLTIESPNEULAGEVGGGAGSGEFEKRDITPTGRTDGAOGGKTSSVTIPELDNKA 1717
QY 56 ERWSILRSVNAVM-----SLADKLGIASSNSSSSRSADVDSTTATATPPP 103
DB 1718 EEGDILAEICINSAMPKSKHKKPRVKIMDQVQOASASSAPNKNQLDGKKKPTSPVKP 1777
QY 104 PPTSDDYKTOAQATYDT-----IFT-----STSLADIQAALVSLQDAV----- 141
DB 1778 IPQNTYTRVRYKNADSKNNLNAERVFSDNKKDSKQNLKNNSKDFNDKLPNEDVRGSGF 1837
QY 142 -----TNIKDT--AATDETAIAAEWETKNADAIVCAQITELAKYASNDQALDLSL 191
DB 1838 AFDSPPHYTPIEGTPYCFSRNDSLSLDDDDDLREKAE-----RKAKENK-----ESE 1890
QY 192 GKLTFSFLLQTLQSVANNKAAELKEM-----QDNPPVPGKTPATAQSLVDQDTATAT 247
DB 1891 AKVTS-----HTELTSNQOASANKTOIAKQPINRGOPKPILO-KOSTFPQSSKDIPDRGAA 1945
QY 248 QIEKDGNAIGDAYFAGONASGAVENAKNSNISIDSAKAAIAATAKTOIAEAKKFPDSP 307
DB 1946 TDEKLQN-----FAIENTPVCFSHNSLSLSLDID----- 1975
QY 308 ILQAEQVMVIOAEKDLKNIKPADGSDVNPCTTVGGSKQ-----QGSSIGSIR 355
DB 1976 --QENNN--KENEPKETETPPDSQGEPSKQASGYAPKSFHVEDTPVCFSRNLSLSLS 2030
QY 356 VSMULDADENETASILMSGFRQIMHFNENTNPDSQAAQQLAAQARAAGDSDSAAL 415
DB 2031 I-----DSEDDLQECISS-----AMPKKKKPSRLKGNKHSPPNMGILGED-LTLDL 2079
QY 416 ADAQKA-LEAALG-----KAGQO--QGILNALGQIASAAYS 449

Db 2080 KDIQRPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAAAAACLS 2124

RESULT 9

US-08-289-548A-7

Sequence 7, Application US/08289548A

Patent No. 5648212

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 102

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner & Allegretti, LTD

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/289,548A

FILING DATE: 12-AUG-1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A.

REGISTRATION NUMBER: 32,141

REFERENCE/DOCKET NUMBER: 1107.46943

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Homo sapiens

IMMEDIATE SOURCE:

CLONE: APC

US-08-289-548A-7

Query Match 6.1%; Score 138; DB 1; Length 2842;

Best Local Similarity 20.0%; Pred. No. 0.11;

Matches 105; Conservative 82; Mismatches 194; Indels 144; Gaps 23;

Qy 11 DETERTPADLSA-OGLEASANKSAEAQRTAGAEAPKE-----SKTDSV----- 55

Db 1658 DLTIESPPELAAAGVGGAQSGEFERDPIETGRSTDQAQGGKTSVVIPELDONKA 1717

Qy 56 ERWSTLRSAVNALM-----SLADKLGIASSNSSSTSRSDVDTTATPTPP 103

Db 1718 EEGDILAEICINSAMPKGSKHPKRVKIMDQVQQAASSAPNQLDGGKKKPTSPVKP 1777

Qy 104 PPTSDDYKTAQATYDT-----IFT-----STSLADIAQALVSLQDAV----- 141

Db 1778 IPONTEYTRVRKNADSKNNLNAERVFSDNKSKQNLKNNKSKDFNOKLPNNEORVRGSGF 1837

Qy 142 -----TNIKDT--AATDEETAAAEWETKNADAIVKGAQITELAKYASDNQALDLSL 191

Db 1838 AFDSPHHYPIEGTPYCFRNSDSLDFDDDDVLSRKAEL-----RKAKNK-----ESE 1890

Qy 192 GKLTSPDLLQTTALLQSVANNKAAELLKEM---QDNVVPVPGKTPAIAQSLVDOTDATAT 247

Db 1891 AKVTS---HTELTSNQOSANKTQATAKOPINRGQPKILO-KOSTFPOSSKDIPDRGAA 1945

Qy 248 QIEKDGNAIGDAYFACQNASGAVENAKSNNSINISDAKAAATATQTAQAEQKKFDPSP 307

Db 1946 TDEKLQN-----FAIENTPVCFSHNSLSLSDID----- 1975

Qy 308 ILQEAQOMVIOAEKDLKNIKPADGSDVPNPVGTTVGSGKQ-----QGSSIGSIR 355

Db 1976 --QENNN---KENEPKETEPDPSQGEPSKPOASGVAPKSFHVEDTPVCFSRNSLSLSL 2030

Qy 356 VSMLLDDAENETASILMSGFRQIMHNTENPDQSQAQOELAAQAAKAAAGDSDSAAAL 415

Db 2031 I-----DSEDDLQECISS-----AMPKKKKPSRLKGDNEKHSRNNMGILGED-LTLDL 2079

Qy 416 ADAQKA-LEAALG-----KAGQO--OGILNALGOIASAAVVS 449

Db 2080 KDIQRPDSEHGLSPDSENFWDKAIQEGANSIVSSLHQAAAAACLS 2124

RESULT 10

US-08-452-654-7

Sequence 7, Application US/08452654

Patent No. 5691454

GENERAL INFORMATION:

APPLICANT: ALBERTSEN, HANS

APPLICANT: ANAND, RAKESH

APPLICANT: CARLSON, MARY

APPLICANT: GRODEN, JOANNA

APPLICANT: HEDGE, PHILIP J.

APPLICANT: JOSLYN, GEOFF

APPLICANT: KINZLER, KENNETH

APPLICANT: MARKHAM, ALEXANDER F.

APPLICANT: NAKAMURA, YUSUKE

APPLICANT: THLIVERIS, ANDREW

TITLE OF INVENTION: INHERITED AND SOMATIC MUTATIONS OF APC

TITLE OF INVENTION: GENE IN COLORECTAL CANCER IN HUMANS

NUMBER OF SEQUENCES: 94

CORRESPONDENCE ADDRESS:

ADDRESSEE: Banner, Birch, McKie & Beckett

STREET: 1001 G Street, NW

CITY: Washington

STATE: D.C.

COUNTRY: USA

ZIP: 20001-4598

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patent In Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/452,654

FILING DATE: 25-MAY-1995

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/741,940

FILING DATE: 08-AUG-1991

ATTORNEY/AGENT INFORMATION:

NAME: Kagan, Sarah A. 32,141

REGISTRATION NUMBER: 1107.035574

REFERENCE/DOCKET NUMBER: 1107.035574

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-508-9100

TELEFAX: 202-508-9299

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:

LENGTH: 2842 amino acids